Bayesian combination of ambiguous shape cues

Wendy J. Adams
Department of Psychology, University of Southampton, Southampton, UK

Pascal Mamassian
Department of Psychology, University of Glasgow, Glasgow, Scotland, UK

We investigate how different depth cues are combined when one cue is ambiguous. Convex and concave surfaces produce similar texture projections at large viewing distances. Our study considered unambiguous disparity information and its combination with ambiguous texture information. Specifically, we asked whether disparity and texture were processed separately, before linear combination of shape estimates, or jointly, such that disparity disambiguated the texture information.

Vertical ridges of various depths were presented stereoscopically. Their texture was consistent (in terms of maximum likelihood) with both a convex and a concave ridge. Disparity was consistent with either a convex or concave ridge. In a separate experiment the stimuli were defined solely by texture (monocular viewing).

Under monocular viewing observers consistently reported the convex interpretation of the texture cue. However, in stereoscopic stimuli, texture information modulated shape from disparity in a way inconsistent with simple linear combination. When disparity indicated a concave surface, a texture pattern perceived as highly convex when viewed monocularly caused the stimulus to appear more concave than a “flat” texture pattern. Our data confirm that different cues can disambiguate each other. Data from both experiments are well modeled by a Bayesian approach incorporating a prior for convexity.

Keywords: cue-combination, depth, texture, stereo, disparity, Bayes, Bayesian

Introduction

The issue of how information about the structure of objects is combined from various sources is an interesting one that has received wide attention. There is much evidence to suggest that different types of visual information are processed in different parts of the brain (Zeki, 1978; Livingstone & Hubel, 1988). An interesting question is, therefore, to what extent different cues influence each other in recovering object properties. Here we aim to distinguish between three different types of cue-combination models using stimuli defined by texture and binocular disparity. In the first model, estimates of shape from texture and from disparity are recovered independently and then these shape estimates are combined. This is “weak fusion” (Clark & Yuille, 1990) or the “weak observer” as described by Landy, Maloney, Johnston, and Young (1995). This is essentially the type of model that has successfully been used to describe cue combination in a variety of studies. Estimates of shape are recovered from independent modules and these estimates are then combined by an averaging process. The weights given to each cue may be determined by the cues’ relative reliabilities (Jacobs, 1999; Backus & Banks, 1999; Ernst & Banks, 2002; Hillis, Ernst, Banks, & Landy 2002).

In the second class of models, the two different cues interact more strongly, such that one cue can influence the interpretation of the other. The “strong observer” lies at the extreme end of cue interaction, whereby depth computation is not necessarily divided into separate modules for individual cues. Such an approach is consistent with the framework proposed by Nakayama and Shimojo (1992). The modified weak fusion (MWF) model, proposed by Landy et al. (1995), lies somewhere between the two, suggesting that only limited interactions occur between cues in an otherwise essentially modular set-up.

In this study we are interested in the stage at which an unambiguous shape cue (disparity) is combined with an ambiguous shape cue (texture). The information contained within the texture pattern of an image is complex because it is probabilistic in nature and relies on making assumptions about the original pattern of texture on the object’s surface. To recover shape from texture, the visual system appears to assume that surface textures are isotropic and homogenous (e.g., Knill, 1998). With these kinds of assumptions, it is possible to infer what surface shapes are more or less likely, given a particular image.

Even with rigid assumptions about the surface texture, at long viewing distances, shape from texture is vulnerable to a sign ambiguity; different shapes give rise to similar patterns of texture in an image. When the effects of perspective projection become negligible at large viewing distances, it is impossible to determine the sign of a slant or curvature. Positive and negative slants are indistinguishable, and
convex and concave objects produce the same image pattern.

In contrast to texture, binocular disparity provides a potentially unambiguous cue to shape, as long as it is correctly scaled for viewing distance. Systematic miscalculations of shape from disparity (Johnston, 1991) have been interpreted as arising from an incorrect estimate of viewing distance (e.g., van Damme & Brenner, 1997). An interesting possibility arises when disparity is presented in conjunction with another visual cue such as texture or motion parallax. Because the two cues scale differently with depth, their combination can, in theory, provide a better estimate of viewing distance (Richards, 1985; Johnston, Cumming, & Landy, 1994; Frisby et al., 1995).

Whether or not disparity is mis-scaled by the incorrect viewing distance, it can still provide adequate information to solve the sign ambiguity in the texture cue. In this work, we are interested in, first, how an ambiguous texture cue is interpreted in the absence of other cues, and second, how that ambiguous cue is combined with binocular disparity information. Finally, we propose a simple Bayesian model that provides a good account of how texture information is used in isolation and in combination with disparity. Our model demonstrates how one cue can effectively disambiguate another, without employing any additional assumptions or distinct stages to do so.

Methods

Observers

Five experienced observers took part in the study. Four were naïve to the purpose of the study and the fifth was an author (WJA). All observers had good stereoaucity.

Stimuli

Stimuli were created using Matlab (MathWorks, Natick, MA). First, a planar texture of randomly distributed lines with randomly selected orientations was created. This surface was then “wrapped” around a vertically oriented ridge. The depth of the ridge varied from –7.5 cm to +7.5 cm. The cross sections of all the ridges were scaled versions of each other and were defined as a portion of an ellipse. The maximum gradient of the largest ridge (12 cm) considered by our model was constrained such that no part of the surface was occluded by itself. Each ridge was positioned in space such that the left and right edges were at the depth of the image plane (164 cm from fixation).

The positions and orientations of the texture elements in the image were generated by projecting them from the surface onto the image plane along the line of sight for the cyclopean eye. The length of all the lines on the screen was 2 mm, thus the compression of texture elements was not a valid cue in our stimuli. This was done to maintain the convex/concave ambiguity. If the texture likelihood is calculated (see model, below) for one of these stimuli, then a bimodal distribution is obtained. The two peaks show that the stimulus is more or less compatible with two interpretations, one convex and one concave with roughly equal depth magnitudes. However, for the purposes of our study, we wanted to create stimuli whose textures were equally compatible with a convex or concave ridge. To ensure this, we randomly selected equal numbers of texture elements from ridges with equal and opposite depths (e.g., 100 texels from a +5 cm ridge and 100 texels from a –5 cm ridge). The texture likelihoods were calculated for these composite stimuli and the random sampling process was repeated until the two peaks of the resultant likelihood were approximately equal (within 10%). These likelihoods are shown in Figure 1.

To create the appropriate disparities for our “texture and disparity” cue-conflict stimuli, each texture pattern was projected back (along the cyclopean line of sight) onto the 3D ridge surface at the appropriate depth. The left and right eyes’ portions of the stimuli were then created separately by projecting the texture elements onto the image plane along the line of sight for the left or right eye. In this way stimuli were generated with disparities defining a particular ridge depth, with a texture that could be consistent with a different depth (Figure 2).

The stimuli extended 6 cm x 6 cm horizontally and vertically and each contained 200 texture lines. Frontoparallel
panels at the top and bottom of the stimuli formed an occluder (Figure 2) that removed the potential of these edges being used as a depth cue. The depth between the occluder and curved textured stimulus was varied randomly from trial to trial.

Stimuli were presented as white lines on a black background using the PsychToolbox for Matlab (Brainard, 1997; Pelli, 1997). The viewing distance was 164 cm. Stimuli were presented on a 21” Sony Trinitron monitor via an arrangement of mirrors forming a modified Wheatstone stereoscope. In the “texture only” condition, observers wore an eye patch over their left eye.

Procedure

On each trial, the stimulus was presented for 2 s. This was followed by the binocular presentation of a frontoparallel contour representing a possible cross section of the stimulus (Figure 3). The initial curvature of the contour was selected randomly on each trial. Using key presses, the observer adjusted the shape of this cross-section line until it matched their perception of the shape of the textured ridge, as if viewed from above. During this adjustment process, the observer had the opportunity to switch between the ridge and the cross section as many times as required until satisfied with their setting. At this point, the screen went blank for 2 s, before a new stimulus appeared.

In the “texture and disparity” condition, each of the four textures (0, ±2.5, ±5, and ±7.5 cm) was presented with each of the seven disparities (–7.5, –5, –2.5, 0, 2.5, 5, and 7.5 cm). In each of three blocks, there were two repetitions of each stimulus (in random sequence) creating a total of 6 responses for each stimulus. In the “texture only” condition, just one eye’s view was presented once for each of the 28 different stimuli in a single block. Results for the “texture only” condition were then averaged over the seven different disparity projections (these minor deviations made no difference to the observers’ settings in the monocular condition).

As a control, two of the observers (WJA and EWG) repeated the experiments with the “texture only” (monocular) and “texture and disparity” (binocular) trials intermingled. In this case the eye to which the monocular stimulus was presented was also varied and no eye patch was worn. Their data (not shown) did not differ significantly from the data when the monocular and binocular trials were presented in separate blocks.

Results

Figure 4 shows the results for the “texture only” stimuli, averaged across the five observers. Error bars give ±1 SE of the mean. The abscissa gives the texture-specified depth and the ordinate shows the perceived ridge depth (the depth difference between the edges and peak of the ellipse,
as indicated by the observers’ cross-sectional settings). The most important aspect of the data to note is that the means lie above zero. In fact, in all trials the stimuli appeared convex (with the exception of the 0-cm texture condition). In other words, the observers discounted the concave interpretation. This is consistent with a prior for convexity that has been noted previously (Mamassian & Landy, 1998; Langer & Bulthoff, 2001; Li & Zaidi, 2001). The second point is that the depth of the ridge was underestimated. This is consistent with a prior for fronto-parallel, and/or the effect of residual cues associated with using a flat monitor, such as accommodation, vergence, and blur cues (Watt, Banks, Ernst, & Zumer 2002; Watt, Akeley, & Banks, 2003).

Figure 5 shows the results for the “texture and disparity” stimuli, again averaged across the five observers. The horizontal axis gives the disparity-specified depth of the ridge, while each line shows the data for one of the four texture conditions. Here again, depth is consistently underestimated. It is possible that depth from disparity is flattened due to an underestimation of viewing distance. However, in our display, all retinal and extraretinal information was consistent with the viewing distance of 164 cm. Therefore, we account for the underestimation of depth in both the “texture only” and the “texture and disparity” experiments as resulting from residual cues to flatness in the display and/or a prior for fronto-parallel (see “Model”). Depth from the “texture and disparity” stimuli is not underestimated to the same extent as depth from texture alone. This is because in the former case there is more shape information available in the stimulus, and, thus, any priors to flatness and/or residual cues have less influence.

Our data pattern suggests that texture is not used in conjunction with disparity to recover a new (incorrect) view at 164 cm. However, in our display, all retinal and extraretinal information was consistent with the viewing distance of 164 cm. Therefore, we account for the underestimation of depth in both the “texture only” and the “texture and disparity” experiments as resulting from residual cues to flatness in the display and/or a prior for fronto-parallel (see “Model”). Depth from the “texture and disparity” stimuli is not underestimated to the same extent as depth from texture alone. This is because in the former case there is more shape information available in the stimulus, and, thus, any priors to flatness and/or residual cues have less influence.

Our data pattern suggests that texture is not used in conjunction with disparity to recover a new (incorrect) viewing distance; this would result in some vastly overestimated and some vastly underestimated depth judgments. A previous study using disparity and texture specified depth also failed to find such rescaling (Frisby et al., 1995).

The positive slope of the lines shows the clear effect of disparity – as disparity-specified depth increased, so did perceived ridge depth. This effect was significant as a main effect in an ANOVA (F(6, 24) = 37.5, p < .01). What we are more interested in, however, is the interaction between disparity and texture. We want to know how a texture cue, which is always interpreted as convex in the “texture only” condition, will be combined with a disparity cue, which signals a convex or concave surface. Consider the solid light blue line (texture = ±7.5 cm) in Figure 5 and compare that to the dark blue dotted line (texture = 0 cm). On the right hand side of the plot, the solid line is above the dotted line, showing that the ±7.5 texture made the stimuli appear to be more convex than the 0-cm texture. This is reasonably straight-forward: a texture that was seen as convex when viewed alone, makes a disparity-specified ridge appear more convex compared to the effect of a flat texture. This is consistent with the data of Buckley and Frisby (1993) for texture and disparity cue combination in convex ridges.

However, a more complex situation arises when disparity signaled a concave surface (left half of the plot). Here, the solid light blue line is below the dotted dark blue one. This means that the ±7.5-cm texture made the surface appear more concave, despite the fact that this texture was seen as convex when viewed in isolation from disparity. This is inconsistent with a linear combination of independent cues. An intuitive way to think of this result is that the concave interpretation appears to be discarded or over-

![Figure 4](https://example.com/figure4.png)

**Figure 4.** Observers’ mean responses (N = 5) for the “texture only” condition. The horizontal axis shows the absolute texture-specified depth of the ridge. Observers’ perceived depth is given on the vertical axis. Error bars are ±1 SEM.

![Figure 5](https://example.com/figure5.png)

**Figure 5.** Observers’ mean responses (N = 5) for the “disparity and texture” condition. The horizontal axis gives the disparity-specified depth of the ridges. Perceived depth is plotted on the vertical axis. Each texture is shown by a different colored line. Error bars are ±1 SEM.
Bayesian model

The Bayesian framework provides an optimal way of combining the information contained within an image with prior assumptions about the nature of objects in the world. This approach has successfully been used to model human behavior in a range of visual tasks. In the current experiment, the sources of information in the image are the disparity and texture cues to shape. These are combined with a prior for convexity and a prior for frontoparallel. A detailed description of our model is provided in the "Appendix." The model is similar in spirit to that presented in van Ee, Adams, and Mamassian (2003).

Texture information can only be exploited by making assumptions about the original texture distribution on the surface in the world. In our model we assume that the distribution of lines is homogenous over the surface (lines are equally likely to be present at any point on the original surface). However, the shape of a ridge means that at different points on the ridge, a given patch size on the surface projects to differently sized patches in the image. This results in systematic changes in texture density across the image: for ridges with large depths, the left and right sides of the resulting image will have a higher density than in the middle of the image (Figure 2). At long viewing distances, where the effects of perspective projection are small, this is largely a function of the local surface slant (see "Appendix"). Similarly, the orientation of a texture line in the image can be calculated from the line's orientation on the original surface, its position, and the local slant of the surface. Because we assume a uniform distribution of texture line orientations on the original surface (the isotropy assumption), the orientation of lines in the image contain information about the probable shape of the object. For example, as the local slant of the surface gets larger, the projection of the texture lines becomes closer to vertical; this can be seen at the left and right edges of the images in Figure 1. Given these assumptions of homogeneity and isotropy, we can calculate the likelihood for any ridge depth, of a line in the image at a particular position with a particular orientation. By considering each line in the image independently and multiplying together these likelihoods for each line, we can calculate the overall perspective likelihood for any image. The perspective likelihoods for the four textures used are shown in Figure 2. There are no free parameters in our model for determining the texture likelihood. Any internal noise is assumed here to be negligible in the context of interpreting a randomly generated texture, whose information content is limited.

The second distribution is the disparity likelihood. This is simply defined as a Gaussian centered on the correct ridge depth. The width of the Gaussian is the first free parameter of the model and reflects the internal noise of the observer. This component of the model is eliminated when the "texture only" stimuli are considered. We must entertain the possibility that errors exist in estimates of depth from stereo due to mis-scaling retinal disparity with the incorrect viewing distance. In our set-up, there were multiple sources of information (vergence, accommodation, known distance to screen, and vertical disparities), all consistent with our viewing distance of 164 cm. We, therefore, chose not to incorporate separate biases into the disparity and texture likelihoods. Our observers' depth judgments in both experiments were well modeled by incorporating a single prior for frontoparallel and/or residual flatness cues.

The third distribution is a prior for convexity. There is varied evidence that in the absence of other information, the visual system "assumes" a convex rather than a concave shape. We have implemented this by using a Gaussian centered on the ridge depth of 3 cm, corresponding to half of the ridge width. In other words, the prior assumption here is for a near-circular cylindrical shape. The spread of this Gaussian is a second, free parameter and reflects the strength of the prior assumption.

Finally, the fourth distribution is a prior for frontoparallel. In limited cue situations, depth is often underestimated. This has been interpreted as reflecting a prior for flatness and/or the presence of residual cues, such a accommodation and blur that arise from using a flat screen to present visual stimuli. The final distribution in our model incorporates both this possible prior and any residual information. It is modeled as a Gaussian centered on zero depth. The width of the distribution (the third and final free parameter) reflects the relative strength of the prior and the reliability of the residual cues.

All of the information – the likelihoods and the priors – are combined by multiplication. This is the optimal combination rule within a Bayesian framework and results in the posterior distribution. It is in this multiplication of the two likelihoods that disparity essentially serves to disambiguate the texture information. For example, consider the case when the texture cue corresponds to a ridge with a depth of ±7.5 cm and the stereo cue corresponds to a ridge depth of –7.5 cm. The texture cue is ambiguous, and its likelihood distribution has peaks at both –7.5 cm and +7.5 cm (see Figure 1). In contrast, the stereo cue is not ambiguous, and its likelihood distribution has only one peak at –7.5 cm. Their product will have a single peak, located at –7.5 cm. In this sense, the stereo cue has disambiguated the texture cue.
We considered two decision rules, one where the output of the model is the maximum of the posterior distribution (MAP) and one where the response is the mean of the posterior. These are equivalent to having either very narrow or very broad gain functions, but in this instance produce very similar results. Maloney (2002) provides an analysis of gain functions. The presented fits from the model were calculated using the mean of the posterior distribution.

Figure 6 shows the individual observers’ data and the best fit from our model. It can be seen that the model provides a good fit for both of the stimulus conditions. For each observer we found the single set of three parameters that provided the best fit (least squared error) to the texture and disparity data and the texture only data. These are given in Table 1. The model’s predictions for the “texture and disparity” data show a kink at around 0 cm disparity specified depth (e.g., observer ML in the ±7.5-cm texture condition). This corresponds to the point at which the peak in the posterior distribution moves from being close to the “concave” peak in the bimodal texture likelihood to being closer to the “convex” peak.

Table 1. Table of fitted parameters to the model.

<table>
<thead>
<tr>
<th>Observer</th>
<th>Disparity Std Dev</th>
<th>Convexity Std Dev</th>
<th>Residual Std Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>WJA</td>
<td>0.378</td>
<td>2.96</td>
<td>0.404</td>
</tr>
<tr>
<td>EWG</td>
<td>0.266</td>
<td>2.07</td>
<td>0.303</td>
</tr>
<tr>
<td>PAW</td>
<td>0.205</td>
<td>0.415</td>
<td>0.292</td>
</tr>
<tr>
<td>LW</td>
<td>1.85</td>
<td>100</td>
<td>0.326</td>
</tr>
<tr>
<td>ML</td>
<td>0.472</td>
<td>1.31</td>
<td>0.357</td>
</tr>
</tbody>
</table>

Table 1. Table of fitted parameters to the model.

**Discussion**

We were interested in examining how ambiguous texture information about shape is interpreted. We also wanted to explore how this cue was combined with disparity, which does not contain the same convex/concave ambiguity. The model that we present here accounts for observers’ behavior, both when texture is presented alone and in conjunction with disparity information. The Bayesian approach that we have used is the ideal way to combine information in the image with prior assumptions. By implementing a prior for convexity, we can capture observers’ behavior when texture is presented alone. However, the same model also describes how texture and disparity are combined. Our observers did not combine texture and disparity in a way that could be described by a simple linear weighting of independent cues. Instead, the disparity information in the stimulus affected which interpretation of the texture information (convex or concave) became dominant. This is predicted in a straightforward way by our model.

This type of disambiguation has also been observed with structure from motion (SFM). Similarly to texture, SFM is prone to a reflection ambiguity, but this ambiguity is resolved by the addition of other cues, such as occlusion and disparity (Braunstein, Anderson, & Riefer, 1982; Proffitt, Bertenthal, & Roberts, 1984; Dosher, Sperling, & Wurst, 1986).

The MWF model (Landy et al., 1995) also provides an explanation for the interaction between cues to resolve ambiguities. Their model involves an explicit “cue promotion” stage, where ambiguities like that in our texture cue would be resolved by other cues. Our model is similar in approach, but we incorporate prior information and have no explicit promotion stage.

**Appendix**

Here we provide details of the Bayesian model used to account for our observers’ behavior in the two stimulus conditions.

**A. Texture likelihood**

In terms of calculations, the texture likelihood can be divided into two components, one related to the likelihood of getting the observed positions of the texture elements in the image, and one component reflecting the likelihood of observing the particular texture line orientations in the image (cf., Knill, 1998).

The ellipse cross section is given by

\[
\frac{x^2}{a^2} + \frac{(z_L - z_0)^2}{b^2} = 1 ,
\]

Figure 6. Individual observers’ data and the best fit of the model. Each column gives a single observer’s data. Each row gives a different texture, except the bottom row, which gives the results for all image textures for the “texture only” condition. Observers’ data are given by circles (“disparity and texture” condition) and stars (“texture only” condition). Each texture is depicted by a different color. The best fit of the model is plotted with black lines. Error bars (±1 SEM) are usually smaller than the symbols.
where \(x_s\) is horizontal position and \(z_s\) is depth on the surface. \(a, b, \) and \(z_o\) are constants that give the maximum horizontal and depth extents of the ellipse and the distance of the center of the ellipse from the image plane, respectively. The center of the ellipse is offset from the image plane; the ridge comprises less than half of a full ellipse. The exact values of \(a, b,\) and \(z_0\) depend on the ridge depth \((h)\).

For a range of ridge depths between –15 and 15 cm, the image is split up into small squares. For each square, the positions on the ridge surface that project to this image patch (located at \(x_i, y_i\)) are calculated.

\[
\begin{bmatrix}
  x_s \\
  y_s \\
  z_s
\end{bmatrix} = \begin{bmatrix}
  x_i(z_s + d) \\
  y_i(z_s + d) \\
  z_0 + b \sqrt{1 - \left(\frac{x_i^2}{a^2}\right)^2}
\end{bmatrix}
\]

The arc lengths on the surface corresponding to the top and bottom sides of a square image patch are calculated by integrating the differential of the curve between the relevant surface points (e.g., \(x_s1\) and \(x_s2\)).

\[
\int_{x_s1}^{x_s2} \sqrt{1 + \left(\frac{b^2 x_s^2}{a^4 - a^2 x_s^2}\right)} \, dx
\]

It is then straightforward to calculate the area on the surface that projects to the image patch under consideration. The probability of a texel lying within this image patch is proportional to the calculated surface area. It can be seen from Figure 2 that for deep ridges, areas of the image near the edges contain, on average, more texels than areas in the middle of the image. Therefore, a texture element at the edge of an image is more likely to have arisen from a ridge with a large depth (convex or concave), whereas a texture element near the middle of an image is more likely to have arisen from a flatter ridge. By doing these calculations for a large range of ridge depths, the likelihood distribution of a particular texel position is given by the probability of this location for the whole range of depths. This position distribution is then multiplied by an orientation likelihood distribution.

The orientation likelihood is calculated, again, by splitting the image into small squares. As with the area calculations, for a large range of ridge depths, the surface patch that projects to each image patch is determined. The slant of the surface \((\phi)\) at that location is given by

\[
\tan(\phi) = \frac{-x_i b^2}{a^2(z_s - z_0)}
\]

The orientation, \(\gamma\), of an image line is then defined as

\[
\tan(\gamma) = \frac{\tan(\phi)(z_s + d) - y_s \sin(\phi)}{\cos(\phi)(z_s + d) - x_s \sin(\phi)}
\]

where \(\theta\) is the orientation of the line on the surface, \(d\) is viewing distance, and \(x_s, y_s\) and \(z_s\) give the center of the line on the surface.

From this equation it is straightforward to calculate the range of \(\theta\) (on the surface) that corresponds to a particular range of \(\gamma\) (in the image). Because the model assumes that \(\theta\) follows a uniform distribution on the original surface, the probability of finding an image line within the range of \(\gamma\) is proportional to the size of the corresponding range of \(\theta\) (Figure 7). The orientation likelihood for any particular image line is found by extracting the probability of that orientation, \(\gamma\), for the complete range of ridge depths. The overall likelihood \(p(t|h)\) for an observed image texture pattern \((t)\) is calculated by multiplying together all of the individual likelihoods for position and orientation for all texture elements.

**B. Disparity likelihood**

The binocular disparity likelihood is modeled as a Gaussian centered on the true disparity-specified ridge depth \((h_b)\). The spread of the distribution \((\sigma_b)\) is left as a free parameter:

\[
p(b|h; \sigma_b) = \frac{1}{\sqrt{2\pi}\sigma_b} \exp\left(-\frac{(h - h_b)^2}{2\sigma_b^2}\right).
\]
C. Convexity prior

A preference, or prior assumption for convex objects, was modeled using a Gaussian centered on a ridge depth of 3 cm (half of the stimulus width, thus close to a circular cylinder). The spread of this distribution, $\sigma_c$, is a free parameter reflecting the strength of the convexity prior:

$$p(h;\sigma_c) = \frac{1}{\sqrt{2\pi\sigma_c^2}} \exp \left[ \frac{-(h-3)^2}{2\sigma_c^2} \right].$$

D. Residual flatness cues

Cues to flatness, arising from stimulus presentation on a flat monitor, along with a possible prior for frontoparallel surfaces are combined in a single Gaussian distribution centered on zero depth. The spread of this distribution ($\sigma_r$) is a free parameter relating to the reliability of the residual cues and the strength of the prior for frontoparallel:

$$p(h;\sigma_r) = \frac{1}{\sqrt{2\pi\sigma_r^2}} \exp \left[ \frac{-h^2}{2\sigma_r^2} \right].$$

E. Combination of likelihoods and priors

The binocular disparity information ($h$) and the texture information in the image ($t$) are combined with the residual cues and prior(s) to produce the posterior distribution. In the “texture only” condition, the disparity likelihood is omitted. The posterior gives the probability of a scene parameter (here ridge depth) given the available image information and prior assumptions. From Bayes’ rule,

$$p(S \mid I) \propto p(I \mid S) p(S).$$

Following the assertion that the disparity and texture cues are independent, the expression becomes

$$p(h \mid t, b; \sigma_b, \sigma_r, \sigma_c) \propto p(t \mid h) p(b \mid h; \sigma_b) p(h; \sigma_c) p(h; \sigma_r).$$

In our model, a response is extracted from the posterior distribution by calculating its mean.

References


Acknowledgments

Thanks to Robert Jacobs for helpful comments. WJA and PM are supported by the Wellcome Trust (grant GR069717MA).

Commercial relationships: none.

Corresponding author: Wendy J. Adams.

Email: w.adams@soton.ac.uk.

Address: Department of Psychology, University of Southampton, Southampton, S017 1BJ, UK.


