This is the peer reviewed version of the following article:
Mitsudo, H., Sakai, A., & Kaneko, H. (2013). Vertical size disparity and the correction of stereo correspondence. *Perception, 42(4),* 385-400,
which has been published in final form at http://dx.doi.org/10.1068/p7387.

Running head: vertical size disparity

Vertical size disparity and the correction of stereo correspondence

Hiroyuki Mitsudo[1*], Ayumi Sakai[2], and Hirohiko Kaneko[3]

[1] Department of Human Sciences, Faculty of Human-Environment Studies, Kyushu University, 6-19-1 Hakozaki, Higashi-ku, Fukuoka 812-8581, Japan. E-mail: hmitsudo@lit.kyushu-u.ac.jp

[2] Graduate School of Human-Environment Studies, Kyushu University, 6-19-1 Hakozaki, Higashi-ku, Fukuoka 812-8581, Japan. E-mail: sakai.ayumi.731@m.kyushu-u.ac.jp

[3] Department of Information Processing, Tokyo Institute of Technology, 4259 Nagatsuta, Midori-ku, Yokohama, Kanagawa 226-8502, Japan. E-mail: kaneko@ip.titech.ac.jp

[*] Corresponding author
Department of Human Sciences, Faculty of Human-Environment Studies, Kyushu University 6-19-1 Hakozaki, Higashi-ku, Fukuoka 812-8581, Japan
E-mail: hmitsudo@lit.kyushu-u.ac.jp
Tel & Fax: + 81926427413

Abstract

We examined the stage of vertical-disparity processing that produces a global stereoscopic slant. In two psychophysical experiments, we measured perceived slant about a vertical axis for two-dimensional stereoscopic patterns consisting of random dots, concentric lines, and radial lines. Binocular image differences were introduced into each pattern by vertically magnifying either the entire image for the right eye or that for the left eye. Because the continuous lines were geometrically ambiguous in local stereo correspondence, the three patterns differed from each other in the local horizontal disparity measured in retinal coordinates. The two experiments revealed that, despite the differences in the retinal horizontal disparity, the slant settings were generally similar for the three patterns, in both short and long viewing distances (25 and 120 cm, respectively). These results are consistent with the idea that the visual system uses vertical disparity at least when establishing local stereo correspondence. A Bayesian model is proposed to account for the results.

Key words: vertical disparity, stereo matching, slant, Bayesian model

1 Introduction

Stereo correspondence refers to the process of analyzing and matching the elements projected onto the two-dimensional (2D) retinae of the two eyes. In natural three-dimensional (3D) scenes, the two retinal images contain slight differences that depend on the distances and directions from the eyes to the viewed objects. The local binocular differences, called horizontal disparities, provide quantitative information about the 3D structure of the scene when extraretinal parameters such as viewing distance and gaze direction are provided (Howard and Rogers 1995, 2012). To gain a better understanding of stereo correspondence, several studies have proposed a variety of theoretical models (Banks et al 2004; DeAngelis et al 1995; Fleet et al 1996).

A challenge to theoretical models of stereo correspondence is the induced effect, a visual phenomenon in which observers perceive a surface slant about a vertical axis when the image presented to one eye is *vertically* magnified with respect to that presented to the other eye (eg Ogle 1950; Serrano-Pedraza and Read 2009). For example, if the entire pattern presented to the left eye is vertically magnified by, say, 5% with respect to that of the right eye, the observer perceives a flat and slanted surface whose left side appears closer. This type of magnification is often called vertical size disparity. The induced effect is typically obtained when the pattern consists of random dots that cover a large portion of the visual field (> 30˚ of visual angle). The perceived slant is approximately proportional to the vertical size disparity within a range of 95-105% (Kaneko and Howard 1996; Ogle 1950).

Because typical stimuli for investigating the induced effect (eg a random-dot or textured pattern) contain zero horizontal disparity across the entire visual field, an additional process must be postulated to explain this phenomenon theoretically. Psychophysical studies have suggested that vertical size disparity is used to correct the global slant of stereoscopic surfaces (Backus et al 1999; Duke and Howard 2012; Kaneko and Howard 1996). A useful formula was proposed by Backus et al (1999), in which horizontal and vertical disparities are represented as horizontal and vertical size ratios, respectively. Let ($x_{0\mathrm{r}}$, $y_{0\mathrm{r}}$) and (x_r, y_r) denote the angular position of two elements presented to the right eye in Cartesian coordinates; let ($x_{0\mathrm{l}}$, $y_{0\mathrm{l}}$) and (x_l, y_l) denote the angular position of the two corresponding elements presented to the left eye. According to Howard and Rogers (2012), the horizontal size ratio H is defined as the ratio of the horizontal angular sizes of two points presented to the two eyes, $H = (x_\mathrm{l} - x_{0\mathrm{l}})/(x_\mathrm{r} - x_{0\mathrm{r}})$. The vertical size ratio V is similarly defined by the ratio of the vertical angular sizes in the two eyes, $V = (y_\mathrm{l} - y_{0\mathrm{l}})/(y_\mathrm{r} - y_{0\mathrm{r}})$. Then, the slant about a vertical axis S (rad) is expressed as a function of H and V:

$$S \approx -\tan^{-1}[\mu^{-1}\ln(H/V)], \qquad (1)$$

where μ is the convergence angle (rad), and ln is the natural logarithm. Under this convention, a positive value of S indicates that the left side of the entire stimulus appears in front of the fixation point. This formula provides a general theoretical account for the induced effect, by representing that the slant suggested by the horizontal-disparity term is corrected (ie divided) by the vertical-disparity term, H/V.

However, the way in which the visual system uses vertical disparity is still unclear. We introduce three possible ideas. First, one can hypothesize that the visual system uses vertical disparity *when and after* establishing stereo correspondence. In this strategy, the visual system can first identify the 2D disparity vector in each spatial position appropriately with the aid of the global pattern of the vertical-disparity field (Backus et al 1999; Duke et al 2006; Kaneko and Howard 1996, 1997; Gårding et al 1995; Matthews et al 2003; Porrill et al 1999). For stimuli that produce the induced effect, the correct direction of correspondence is vertical. After establishing stereo correspondence appropriately, the global vertical disparity is used to interpret horizontal disparities by applying Equation (1). This strategy generally agrees with the conventional explanation of the induced effect and is called the standard theory of the induced effect here.

Second, one can hypothesize that the visual system could use vertical disparity *only after* establishing stereo correspondence. In this strategy, the visual system first identifies the 2D disparity vector at purely *local* level, without the aid of the global vertical-disparity field. Vertical disparities are pooled across different positions and used to correct horizontal disparities. This strategy is called the late-correction hypothesis here.

Third, one can hypothesize that the visual system uses vertical disparity *only when* establishing stereo correspondence (Julesz 1971; Ogle 1950). For example, Ogle (1950) argued that the visual system could first identify the global vertical mismatch between the two retinal images and then cortically enlarge the entire retinal image projected onto one eye so that the overall heights of the two images are equal. Consequently, the overall widths of the images are now different. By processing the horizontal disparities between these corrected images, the visual system produces a global slant. In this paper we call this strategy the early-correction hypothesis.

To examine the three ideas, here we devised two spatial patterns consisting of continuous lines, concentric and radial (Figure 1A). Random-dot patterns contain spatially distributed elements, which provide sufficient visual information to establish local stereo correspondence as well as to measure horizontal disparities unambiguously. Random-dot stimuli are therefore not adequate to examine this issue. In contrast to random-dot stereograms, continuous lines [as well as continuous edges and one-dimensional (1D) gratings] are geometrically ambiguous in local stereo correspondence (Arditi 1982; Farell 1998; van Ee and Schor 2000; Mitsudo 2007). In the concentric pattern, the local orientation of each line is almost perpendicular to the direction of the line position with respect to the center of the stimulus. In the radial pattern, on the other hand, the local orientation of each line is always parallel to the direction of the line position with respect to the stimulus center.

Adding a vertical size disparity to the concentric and radial patterns results in retinal horizontal disparities that depend on both the line orientation and the retinal position. Consider a case where the vertical size ratio b introduced into the whole stimulus is greater than 100% (eg H = 100%; $V = b$ = 105%, Figure 1A). For the random-dot pattern, there is no ambiguity in local stereo correspondence, yielding $H/V = 1/1.05 \approx 0.952$. Throughout this paper we consider a situation where the observer judges the overall slant of the entire stimulus while fixating the center of the stimulus placed in front of the head at a constant viewing distance.[1] Therefore, the local

---

[1] With this definition, the size ratio is not necessarily correlated with the disparity gradient (see Howard and Rogers 1995, for mathematical deduction), especially for the concentric pattern. However, the definition is sufficient for considering our situation

horizontal size ratio is computed with respect to the stimulus center. This means $(x_{0r}, y_{0r}) = (x_{0l}, y_{0l}) = (0, 0)$, simplifying the calculation of the horizontal size ratio, $H = (x_l - x_{0l})/(x_r - x_{0r}) = x_l/x_r$.

The three ideas provide predictions about slant for the three patterns. First, the standard theory states that the visual system can somehow establish correct correspondence for the two line patterns as well as the random-dot pattern, by taking into account the global vertical-disparity field at this correspondence stage. Because the correct direction of correspondence is vertical in the three patterns, the predicted slants for the two line patterns will be identical to that for the random-dot patterns.

Second, let us consider a specific prediction from the late-correction hypothesis. We assume that the visual system prefers horizontal correspondence for the line patterns at local level, as reported in psychophysical studies (Morgan and Castet 1997; van Ee and Schor 2000). This means that the vertical component is always close to zero (V = 100%) for the line patterns. For the concentric pattern, a vertical size ratio of 105% produces a *crossed* retinal horizontal disparity in the first and fourth quadrants (ie $x_l > x_r$; $x_l, x_r > 0$), and an *uncrossed* horizontal disparity in the second and third quadrants (ie $x_l < x_r$; $x_l, x_r < 0$). The values of the local horizontal size ratio H are then greater than 100% (eg approximately 104.5% in a 45° oblique position).[2] Therefore, the values of H/V become greater than one, yielding a slant opposite to that in the random-dot condition. For the radial pattern, on the other hand, introducing the same value of the vertical size ratio b results in a retinal horizontal disparity whose direction is opposite that of the concentric pattern in each quadrant (Figure 1A). The value of H becomes 95.2% ($H = 1/b$), and therefore the value of H/V is 0.952, yielding a slant which is the same as that of the random-dot pattern. The late-correction hypothesis predicts generally different slants for the three patterns even when different assumptions are made. See Appendix A.

Third, we consider the early-correction hypothesis. For each stereogram, the corresponding point in the image of the left eye can be represented by $(x_l, y_l) = (x_r, by_r)$. If the stereo system corrects correspondence by magnifying the entire image of the left eye by $1/b$, the point in the left eye becomes $(x_r/b, y_r)$. See Figure 1B. Then, the value of H is $(x_r/b)/x_r$ = 95.2%, and the value of V is y_r/y_r = 100%, yielding $H/V = 0.952/1 = 0.952$. This value is the same as the prediction from the other hypotheses for the random-dot pattern. Consequently, the prediction from the early-correction hypothesis is identical to that from the standard theory: Predicted slants are generally similar for the three patterns.

---------------------Figure 1 around here ---------------------

In the two experiments reported here, we tested these predictions by measuring the perceived slant about a vertical axis for the random-dot, concentric, and

---

because (a) the disparity gradient is related to the local slant, and (b) we are interested in the overall, global slant.

2 In the concentric pattern, H depends on the spatial position and is represented by $[1 - (b^{-1}\sin t)^2]^{0.5}/\cos t$, where $t = \sin^{-1}[y_r/(x_r^2+y_r^2)^{0.5}]$. Our analysis assumes that the visual system averages local estimates of horizontal disparity measured across different spatial positions to obtain a global slant.

radial patterns as a function of the vertical size ratio, varied from 95.2% to 105% across trials. Using the method of adjustment, the observers reported the overall slant of each pattern at viewing distances of 25 and 120 cm (Experiments 1 and 2, respectively). In Section 4, we tried to construct a simple computational model to account for our psychophysical data—the slant settings were generally similar for the three patterns in Experiments 1 and 2. In the present model, the perceived slant was assumed to derive from horizontal disparities calculated after a correction for the binocular difference in overall image size (ie image magnification). Although Ogle (1950) pointed out that the correction of stereo correspondence could account for the induced effect, no study has examined this idea quantitatively. We adopted a Bayesian framework to determine the amount of image magnification to be corrected from stereo images, as in the case of cyclovergence (a counter-rotation of the two eyes about the line of sight, Mitsudo et al 2009).

2 Experiment 1

2.1 Methods
2.1.1 Observers

Ten observers (aged between 19 and 33 years old) participated in the experiment. Eight were naive as to the purpose of the experiment, and two were the authors. All had normal or corrected-to-normal visual acuity and were able to see the depth defined by horizontal disparity in a simple anaglyph stereogram.

2.1.2 Apparatus

Stereo stimuli were presented on a 15-inch LCD color monitor (Sharp LL-T1502T, 1024 x 768 pixels) with the anaglyph technique. A personal computer (Dell Dimension 8400) was used to control the stimulus presentation and to obtain the observer’s response. One pixel subtended 4 x 4′ of visual angle. All stimuli were drawn with the antialiasing method.

2.1.3 Stimuli

The test stimuli were random-dot, concentric, and radial patterns drawn in red and blue (Figure 1A). Each pattern subtended approximately 30˚ of visual angle in diameter. Vertical size disparity was introduced into each pattern by 95.2%, 97.6%, 100%, 102.5%, and 105%. A fixation with nonius lines (2.7 x 2.7˚ of visual angle) was presented at the center of the screen. The stimulus for the right eye image was drawn in blue and viewed through a blue filter; the stimulus for the left eye was drawn in red and viewed through a red filter. The interocular crosstalk was small (5.5% on average) enough to produce the desired disparity. The background of the screen was dark. The three patterns were matched with each other in terms of the number of pixels that constituted each pattern. This was done by adjusting the number and size of dots for the random-dot pattern and the number, length, and width of lines for the concentric and radial patterns. Consequently, the stimulus parameters were determined as follows. The random-dot pattern consisted of approximately 750 dots, each of which subtended 0.4 x 0.4˚ of visual angle. The concentric pattern consisted of seven ovals; the widths of the largest and smallest ovals were 32˚ and 8˚ of visual angle, respectively, and the line width of each oval was 0.3˚ of visual angle. The radial pattern consisted of 38 line segments; both ends of each line segment were gradually darkened so that they did not produce a clear line termination.

The matching display contained two lines (length, 30˚ of visual angle), one for adjustment and the other as a reference (Figure 2). The continuous line was adjustable and could be rotated by the observer. The step size of adjustment was 0.2˚. The dashed line, whose orientation was fixed and always horizontal, served as a reference representing the fronto-parallel plane for the task. At the beginning of the adjustment for each trial, the continuous line was superimposed on the dashed line at the center of the screen. In addition to these lines, a binocularly correlated, random-dot pattern with zero disparity was presented during the adjustment.

---------------------Figure 2 around here ---------------------

2.1.4 Procedure

The observer binocularly viewed the screen, which was placed in front of the head at a viewing distance of 25 cm. The experiment was conducted in a darkened room; no objects were visible except for the stimulus. The test and matching stimuli were presented at the center of the screen. The observer's head was stabilized with a forehead-and-chin rest.

At the beginning of each trial, the observer was required to view the test pattern for 4 s while fixating the nonius pattern presented at the center of the pattern. The test display then disappeared, and the matching display appeared. The observer's task was to adjust the orientation of the matching line stimulus until it appeared to represent the perceived slant of the test stimulus. The observer was instructed to regard the matching line as the schematic cross-section of the test stimulus with respect to the fronto-parallel plane (ie as if the test stimulus were viewed from above the screen). During the adjustment, the observer was not required to fixate the center of the pattern. The adjustment was made by pressing the "4" and "5" keys on the extended keyboard.

Each block comprised ten trials (five vertical size ratios x two repetitions). During each block, the order of the vertical size ratios was randomized, while the type of test pattern was constant. After several practice blocks, each observer completed nine blocks (three blocks for each test pattern). The presentation order of the test patterns was counterbalanced across blocks and observers.

2.2 Results and discussion

Figures 3A-C show the mean slant settings averaged over the ten observers. The oblique gray line represents the predicted slant based on Equation (1). A two-way repeated-measures analysis of variance (ANOVA) was performed on the mean slant settings, with the factors of vertical size ratio and stimulus pattern. Only the main effect of vertical size ratio was significant, $F(4, 36) = 5.6, p < .005$.

To examine the overall effect of vertical size ratio on perceived slant, we computed the gain of perceived slant as a function of vertical size ratio for each test pattern. In particular, the slant settings were fitted with a modified version of Equation (1),

$$S \approx -g\tan^{-1}[\mu^{-1}\ln(1/V)] + c_0, \qquad (2)$$

where g is the gain of perceived slant as a function of vertical size ratio, and c_0 is constant overall bias in slant settings. The least-squares method was used to estimate the two free-parameters, the gain g and the constant c_0, for each observer. Figure 3D shows the values of gain for the random-dot, concentric, and radial patterns. The mean of gain values was positive. A one-way repeated-measures ANOVA was conducted on the gain value, with the factor of stimulus pattern. No main effect of stimulus pattern was significant, $F(2, 18) = 1.78$, $p = 0.20$.

---------------------Figure 3 around here ----------------------

To summarize, (a) the direction of the obtained slants was consistent with that predicted from Backus et al's (1999) model (Equation 1) for all the three patterns, and (b) the magnitude of the obtained slants was less than the predicted value, but similar among the three patterns tested. These results generally support both the standard theory and the early-correction hypothesis.

In this experiment, the observer binocularly fixated the center of large test stimuli at a short viewing distance, 25 cm. Under this condition, the two retinal images differed considerably from each other in overall shape. For example, the left side of the stimuli viewed from the left eye was larger than that viewed from the right eye, thereby creating vertical and horizontal disparities especially at eccentric retinal positions. This type of information is called differential perspective—spatial gradients in the horizontal and vertical size ratios (Rogers and Bradshaw 1993). In Experiment 1, differential perspective in theory provides information about the observation distance, but not about the slant of a surface. Nevertheless, one might suspect that perceived slants are similar for the three patterns only when the stimulus contains a considerable amount of the differential-perspective cues to observation distance. To examine this issue, we conducted a second experiment.

3 Experiment 2

Experiment 2 employed similar settings to those used in Experiment 1 and examined whether the results of Experiment 1 could be replicated with a longer viewing distance (120 cm). When the stimuli used in Experiment 1 are viewed from this distance, differential perspective becomes very small. If differential perspective plays a minor role in the present experiments, slant settings were expected to be similar for the three test patterns.

3.1 Method

3.1.1 Observers

Five observers (aged between 24 and 46 years old) participated in the experiment. All were naive to the purpose of the experiment, except for two of the authors. One of the observers participated in Experiment 1. All had normal or corrected-to-normal visual acuity and reported that they were able to see stereoscopic depth.

3.1.2 Apparatus, stimuli, and procedure

The stimuli and procedure were essentially identical to those used in Experiment 1, except that the stimuli were displayed on a large, back-projected screen. The observer viewed stereoscopic stimuli with a pair of liquid-shutter stereoscopic glasses (NuVision 60GX). The refresh rate of the screen was 59.9 Hz for each eye. The stimuli were generated by an Apple MacBook Pro. The viewing distance was approximately 120 cm. In Experiment 2, one pixel subtended 8 x 8′ of visual angle, which was larger than 4 x 4′ in Experiment 1. The number of dots was approximately 270 in the random-dot pattern; four ovals were presented in the concentric condition, so that each stimulus pattern subtended approximately 40° of visual angle in diameter.

3.2 Results and discussion

Figures 4A-C show the mean slant settings averaged over the five observers. The oblique gray line represents the predicted slant based on Equation (1). As in Experiment 1, we conducted a two-way repeated-measures ANOVA on the mean slant settings, with the factors of vertical size ratio and stimulus pattern. We found that the two-way interaction was significant, $F(8, 32) = 2.504$, $p < 0.05$, as was the main effect of the vertical size ratio, $F(4, 16) = 10.196$, $p < 0.0005$.

We computed the gain of perceived slant as a function of vertical size ratio as in Experiment 1. Figure 4D shows the values of gain for the three patterns, averaged over the five observers. A one-way repeated-measures ANOVA was conducted on the gain value, with the factor of stimulus pattern. No main effect of stimulus pattern was significant, $F(2, 8) = 2.68$, $p = 0.13$. These gain values were positive and similar to those obtained in Experiment 1 (approximately 0.2).[3] These results suggest that the differential-perspective cues to viewing distance play a minor role in perceiving slant from the line stereograms.

---------------------Figure 4 around here ----------------------

Although the results of Experiment 2 generally replicated those of Experiment 1, the perceived slant tends to be smaller for the concentric pattern than for the random-dot and radial patterns. Indeed, we found a significant interaction between the vertical size ratio and stimulus pattern in Experiment 2, but not in Experiment 1. We currently have no clear explanation for this difference, but can point out two possibilities. First, a relatively smaller slant for the two line patterns may be partly explained by the perspective cues contained by the stimuli (Banks and Backus 1998), because the arrangement of the elements contained by the two line patterns was regular, therefore suggesting the fronto-parallel plane. Second, it may be related to a large individual difference in metric judgments about stereoscopic depth (Harris et al 2012). This issue will be discussed again in Section 4.3.

4 Bayesian model

The results of the two experiments showed that observers perceived similar slants for the three patterns with vertical size disparity, irrespective of the different retinal horizontal disparities. These results are generally consistent with both the

---

[3] We conducted a two-way mixed-design ANOVA on the gain value, with the factors of stimulus pattern (random-dot, concentric, radial) and viewing distance (25 cm, 120 cm). Neither the main effect of stimulus pattern nor that of viewing distance was significant [$F(2,26) = 1.21$, $p = 0.31$; $F(1, 13) = 0.078$, $p = 0.78$, respectively]. The two-way interaction was also not significant, $F(2, 26) = 0.956$, $p = 0.40$. These results confirm that the gain values do not systematically vary across the experimental conditions tested.

standard theory and the early-correction hypothesis. Because the early-correction hypothesis seems to be simpler than the standard theory, we decided to make a computational model based on the early-correction hypothesis.

As discussed in the Introduction, the early-correction hypothesis states that the visual system adjusts the image-size difference when establishing stereo correspondence. This idea is a promising, but not satisfactory quantitative explanation because available visual signals might be insufficient to estimate how much of the image-size correction is needed. In this section, we tried to show that image-size correction is possible by extending a computational framework of stereo correspondence. To construct a quantitative model to account for the psychophysical data, we take two additional factors into account. One is that the visual system should search for corresponding elements within a limited range of magnification and should prefer a correspondence with a smaller amount of magnification. The other is that, when adjusting the image-size difference, the visual system should take into account the constraint that retinal disparities arise not only from magnification but also from the 3D structure of the scene. Retinal disparities contained by the line stereograms can be interpreted as arising from an image-size difference or the local horizontal disparities created by depth changes. We included these two factors in a Bayesian model.

The present model consisted of two stages. In the first stage, a stereo pair was analyzed to estimate the overall difference in size (ie magnification) between the images presented to the two eyes. We adopted and modified a correlation-based model (eg Banks et al 2004; Mitsudo et al 2009), in which local disparity is determined by binocular cross-correlation between small regions (called windows) within the images. We assumed that visual signals, not extraretinal signals regarding eye position, are used to estimate the value of image magnification. This assumption was based on the psychophysical results that visual signals determine the perceived slant of the surfaces presented in various gaze directions when the stimuli are large and 2D (Backus et al 1999). In the second stage, horizontal disparities were determined from the corrected stereo pair to obtain a final 3D shape (Figure 5).

---------------------Figure 5 around here ---------------------

4.1 Assumptions

In the first stage of the model, we computed and assessed the posterior probability of image magnification β for a given stereo pair. According to the Bayesian view, the posterior probability $p_{\mathrm{posterior}}(\beta|I_{\mathrm{l}},I_{\mathrm{r}})$ is proportional to a product of the prior probability $p_{\mathrm{prior}}(\beta)$ and the likelihood of obtaining the stereo pair $L(I_{\mathrm{l}},I_{\mathrm{r}}|\beta)$, where I_{l} and I_{r} are 2D images presented to the left and right eyes, respectively. Similarly in earlier Bayesian models (Mitsudo et al 2009; Weiss et al 2002), the prior probability distribution of image magnification is assumed to be a Gaussian function with a mean of 0% (corresponding to a size ratio of 100%) and a fixed value of standard deviation σ, $p_{\mathrm{prior}}(\beta) = (2\pi\sigma^2)^{-1/2}\exp[-\beta^2/(2\sigma^2)]$. This assumption reflects that large vertical disparities are unlikely to occur in natural scenes. The standard deviation of the prior distribution σ was the free parameter of our model.

In our model, the likelihood of obtaining the stereo pair $L(I_{\mathrm{l}},I_{\mathrm{r}}|\beta)$ is a function of assumed image magnification β. In computing the likelihood, positional correspondence between a stereo pair was analyzed to determine the local disparity d_j at the 2D position (x_j,y_j) on the image plane, where j is an index to the analyzed position. To implement a correction of the global geometry for stereo correspondence, we magnified the image for the left eye by β while holding both the image for the right eye and the epipolar lines constant. See Appendix B for the details of calculating the likelihood.

In the second stage of the model, by finding and using the value of β that maximizes the posterior probability, we determined the disparity map $\boldsymbol{\delta} = \{\delta_1,\delta_2,\delta_3,\ldots,\delta_n\}$, where n is the number of analyzed image positions. The disparity map parallels the 3D percept produced by the stereo pair.

4.2 Stimuli and analysis

The analyzed images were the stereograms used in Experiment 1. We had three stimulus patterns (random-dot, concentric, radial) and five levels of vertical size ratio (95.2, 97.6, 100, 102.5, 105%). The size of each image was scaled down to 350 x 350 pixels. The images were blurred by a 2D Gaussian blur kernel (size, 5 x 5 pixels, corresponding to approximately an s.d. of 1 pixel) to simulate retinal stimulation in the periphery. For each stereo pair, the posterior probability was calculated with the value of β ranging from –10 to 10%.

To calculate the predicted slant about a vertical axis for each stereo pair, we fitted the estimated horizontal disparities to the horizontal positions of matching elements $\mathbf{x} = \{x_1,x_2,x_3,\ldots,x_n\}$. We used a linear regression, $\boldsymbol{\delta} = a_0 + a_1\mathbf{x} + \varepsilon$, where a_0 and a_1 are free parameters, and ε is a random variable. The slope of the linear regression a_1 was used to compute the predicted slant S_{predict} (rad) for each stereo pair by using the equation:

$$S_{\text{predict}} \approx -\tan^{-1}(a_1 D/i), \qquad (3)$$

where i is the assumed interocular distance (6.5 cm), and D is the viewing distance used in Experiment 1 (25 cm). See Appendix C for derivation.

4.3 Results and discussion

Figures 6A-D show the predicted slants as a function of vertical size ratio, obtained with different values of the standard deviations of the prior distribution (σ = 0.01, 0.21, 0.41, and 0.81%, respectively). A smaller value of σ implies a case in which the stereo system tends to ignore vertical disparities. When the value of σ was higher than 0.21%, the direction of the predicted slant for the random-dot pattern was consistent with the geometrical prediction for the induced effect (Equation 1). When the value of σ was higher than 0.41%, predicted slants were similar to each other for the three patterns. This shows that the model can account for the psychophysical data reported in Experiments 1 and 2.

It is notable that the relative magnitude of predicted slants somewhat differed among the three patterns, depending on the standard deviation σ of the prior distribution of image magnification. For example, the predicted slants for the concentric and radial patterns were more similar when the value of σ was 0.81% than when this value was 0.41%. These variations are comparable to small fluctuations in gain g found in Experiments 1 and 2. A variation in the prior distribution may therefore explain individual differences in the perceived slant[4].

---------------------Figure 6 around here ---------------------

5 General discussion

The two experiments showed that, although the three patterns differed in terms of the retinal horizontal disparities, slant settings depended mainly on vertical size disparity and were generally similar for the different patterns. These data support both the standard theory and the early-correction hypothesis, but not the late-correction hypothesis. This implies that the visual system uses vertical disparity *at least when* establishing stereo correspondence. As a possible algorithm for explaining the present data, we proposed a Bayesian model in which the stereo system estimates and uses interocular image magnification.

One might think that the correction of stereo correspondence is unlikely, because a recent study reported psychophysical data showing that image-size differences impaired stereo correspondence (Vlaskamp et al 2009). However, the results reported by Vlaskamp et al (2009) are not necessarily inconsistent with the present data and model because the magnification value that produced an impairment in their task (±10-15%) was greater than the magnification range used in the present study (approximately ±5%).

There is a long-standing debate on whether vertical disparity plays a critical role in the induced effect (Arditi 1982; Howard and Rogers 2012; Matthews et al 2003). Matthews et al (2003), for example, have proposed a parsimonious explanation by hypothesizing that the induced effect arises from receptive-field properties of cortical neurons that encode disparities. Their central assumptions are that (a) a stimulus that typically produces the induced effect (eg a random-dot stereogram) contains all orientation components at a given retinal position, including the radial orientation with respect to the center of the visual field, and (b) binocular neurons in visual areas have a radial bias in their orientation selectivity. If so, at oblique positions in the periphery, neurons sensitive to non-zero horizontal disparities could respond to non-zero vertical disparities, because of the stereo aperture problem. These hypothetical responses would in theory account for the induced effect.

The present results for the concentric pattern provide evidence against this explanation. The concentric pattern used here did not contain the radial orientation component. Therefore, if Matthews et al's explanation is correct, the perceived slant would be zero across the range of vertical size ratios tested. In Experiments 1 and 2, however, we found that the slant settings varied according to the vertical size ratio. Our results therefore do not support Matthews et al's (2003) hypothesis, and suggest that vertical disparity plays a critical role in the induced effect.

---

[4]When σ was larger than 0.81%, predicted slants (data not shown) were very similar to those obtained with $\sigma = 0.81\%$. This was the case when the prior distribution was not used (ie maximum likelihood estimation). Note that the σ range that produces similar slants (ie above 0.41%) also depend on the number of cross-correlations by which the likelihood of image magnification is calculated in the present model. If the number of cross-correlations is smaller, the σ value required to produce similar slants will be higher.

The line stereograms used here produce not only retinal horizontal disparity but also local orientation disparity. Some studies have pointed out that orientation disparity would also provide useful information about surface slant (Koenderink and van Doorn 1976; Greenwald and Knill 2009). The results of the present experiments do not rule out the possibility that orientation disparity contributes to the similarity in the slants obtained for the three patterns. To reveal the role of orientation disparity in the slant seen with the line stereograms, further studies will be necessary.

In the two experiments, the magnitude of perceived slants was consistently smaller than that suggested by the geometrical prediction of the induced effect, even for the random-dot pattern (approximately a gain of 0.2-0.3). (This result is not explained by our model.) Similar results were reported in other psychophysical experiments using a matching task (Duke et al 2006; Gillam et al 1988; Kaneko and Howard 1996). These results contradict other studies using a nulling task, which reported that the magnitude of the induced effect was comparable to that of the geometrical prediction (Backus et al 1999; Ogle 1950). The discrepancy might partly result from the cue-combination strategy used in the visual system (Backus and Banks 1999).

The present model provides a computational and quantitative account for why the perceived slant was similar for the patterns that contained different values of retinal horizontal disparity. Whereas neuronal mechanisms that encode binocular disparity have been investigated extensively, no study has reported on neurons that change the position of their receptive field according to vertical disparity. Further studies will be necessary to reveal the psychophysical and physiological mechanisms in detail.

Appendix A. Different versions of the late-correction hypothesis

We consider two different versions of the late-correction hypothesis and their predictions about slant for the line stereograms. First, we can assume that the visual system follows the nearest-neighborhood rule to establish local stereo correspondence. This rule says that the correspondence direction is approximately perpendicular to the local orientation of each line. For the concentric pattern, if we assume that the correspondence direction is perpendicular to the local orientation of the line presented to the right eye, a pair of corresponding positions satisfies $y_l/x_l = y_r/x_r$. In this case, we obtain $H/V = (x_l/x_r)/(y_l/y_r) = x_l y_r / x_r y_l = x_r y_l / x_r y_l = 1$, suggesting a zero slant for the concentric pattern. For the radial pattern, if we assume that the correspondence direction is perpendicular to the imaginary line that bisects the right and left images of each radial line, the corresponding positions have the same distance R from the fixation (ie retinal eccentricity). Let t_r and t_l represent the angles of the radial line (with respect to the horizontal meridian) for the right and left eyes, respectively. By definition, $\tan t_l = b \tan t_r$. The corresponding position can be represented by ($R\cos t_r$, $R\sin t_r$) and ($R\cos t_l$, $R\sin t_l$) for the right and left eyes, respectively. Given these assumptions, $H/V = (R\cos t_l / R\cos t_r)/(R\sin t_l / R\sin t_r) = \cos t_l \sin t_r / \sin t_l \cos t_r = \tan t_r / \tan t_l = 1/b$. For the radial pattern, the predicted slant is therefore the same as that of the random-dot pattern.

Second, we can assume that the visual system detects horizontal and vertical disparities independently. Specifically, as found in psychophysical studies (Morgan and Castet 1997; van Ee and Schor 2000), the stereo system could extract retinal horizontal disparities from local line elements. In this case, the values of H are the same as those stated in the main text; the global vertical disparity b could be estimated reliably from given images and *independently* of horizontal disparities. (Note that this assumption allows the visual system to establish dual correspondence, horizontal and vertical.) Consequently, for the concentric pattern presented with a vertical size ratio of 105%, the values of H/V (104.5/105) become around one in a 45° oblique position, suggesting a zero slant. For the radial pattern, the value of H/V (95.2/105) becomes 0.907. Therefore, the predicted slant for the radial pattern would be approximately twice as steep as that for the random-dot pattern.

Appendix B. The computation of the likelihood of stereo correspondence

We used and modified the binocular cross-correlation model (Banks et al 2004; Mitsudo et al 2009) to compute the likelihood of obtaining a stereo pair with assumed magnification β and to obtain a 3D shape from a stereo pair. In the model, cross-correlation is computed between small regions (called windows) within the images for the two eyes. We applied the cross-correlation window to the intersections of an imaginary grid with intervals of 17 pixels over almost all the area of the stereo images (diameter, 300 pixels). The value of local disparity was determined by the displacement that yielded the maximum value of cross-correlation:

$$\rho_j(\delta,\omega) = \frac{\sum_{h=-w}^{w}\sum_{k=-w}^{w}\{I_l[(x_j+h)(\beta+1),(y_j+k)(\beta+1)]-\mu_l\}[I_r(x_j+h-\delta,y_j+k-\omega)-\mu_r]}{\sqrt{\sum_{h=-w}^{w}\sum_{k=-w}^{w}\{I_l[(x_j+h)(\beta+1),(y_j+k)(\beta+1)]-\mu_l\}^2}\sqrt{\sum_{h=-w}^{w}\sum_{k=-w}^{w}[I_r(x_j+h-\delta,y_j+k-\omega)-\mu_r]^2}},$$

(B.1)

where j is an index to the window position, w is the parameter that determines the width and height of the window, h and k are integers, δ and ω are the horizontal and vertical displacement of the window for the right eye, and μ_l and μ_r are the mean intensities of the window area in the images for the left and right eyes, respectively. In the analysis presented here, w was set to 4 so that the size of each correlation window was 9 x 9 pixels. [Mitsudo (2012) showed that a change in the window size has little effect on the likelihood.] To implement a correspondence preference for the zero-disparity plane, cross-correlation was multiplied by a 2D Gaussian-like function,

$$f(\delta,\omega) = \exp[-\delta^2/(2c\delta_{\mathrm{range}}^2)]\exp[-\omega^2/(2c\omega_{\mathrm{range}}^2)], \quad \text{(B.2)}$$

where $\delta_{\mathrm{range}} = 15$, $\omega_{\mathrm{range}} = 5$, and $c = 5$. Consequently, local 2D disparity (d_j, v_j) was determined by assessing $\mathrm{argmax}_{\delta,\omega} f(\delta,\omega)\rho_j(\delta,\omega)$. The value of local correlation was then determined by $r_j = f(d_j,v_j)\rho_j(d_j,v_j)$. The constant δ_{range} was the range for finding the local maximal horizontal disparity; a δ_{range} of 15 pixels was equivalent to approximately 1.5° of visual angle in our experimental settings. The constant ω_{range} corresponded to vertical tolerance in local stereo correspondence. With a smaller value of ω_{range}, predicted slants were more similar to each other for the three stimulus patterns (data not shown).

The logarithm of the likelihood of obtaining a stereo pair as a function of magnification was calculated by

$$\ln L(I_l, I_r|\beta) = -n\ln C - (n/2)\ln[(2\pi/n)\textstyle\sum^{n}_{j=1}(r_j - 1)^2] - n/2, \qquad \text{(B.3)}$$

where n is the number of cross-correlations, $C = \int^{1}_{-1}(2\pi s^2)^{-1/2}\exp[-(r-1)^2/(2s^2)]dr$, and $s^2 = (1/n)\sum^{n}_{j=1}(r_j - 1)^2$. See Mitsudo (2012) for derivation.

Appendix C. Determining global slant from a disparity map

Let E represent the horizontal position between the fixation and a given point (Figure C). Then the disparity of the point on the fixation plane is represented by a_1E. When the point has depth d with respect to the fixation plane, $i/(D-d) = a_1E/d$, where d is positive for a point with a crossed disparity. Furthermore, given that slant is equal to the angle between the element and the vertical meridian, $S = -\tan^{-1}(d/E)$. From these two equations, we obtain $S = -\tan^{-1}[a_1D/(i + a_1E)]$. When we consider the slant around the fixation, E should be close to 0, yielding $a_1E \approx 0$. We therefore obtain Equation (3).

---------------------Figure C around here ----------------------

Acknowledgements

This study was supported in part by KAKENHI 20730478. We thank the two anonymous reviewers for their helpful comments.

References

Arditi A, 1982 “The dependence of the induced effect on orientation and a hypothesis concerning disparity computations in general” *Vision Research* 22 247-256

Backus B T, Banks M S, 1999 “Estimator reliability and distance scaling in stereoscopic slant perception” *Perception* 28 217-242

Backus B T, Banks M S, van Ee R, Crowell J A, 1999 “Horizontal and vertical disparity, eye position, and stereoscopic slant perception” *Vision Research* 39 1143-1170

Banks M S, Backus B T, 1998 “Extra-retinal and perspective cues cause the small range of the induced effect” *Vision Research* 38 187-194

Banks M S, Gepshtein S, Landy M S, 2004 “Why is spatial stereoresolution so low?” *The Journal of Neuroscience* 24 2077-2089

DeAngelis G C, Ohzawa I, Freeman R D, 1995 “Neuronal mechanisms underlying stereopsis: how do simple cells in the visual cortex encode binocular disparity?” *Perception* 24 3-31

Duke P A, Howard I P, 2012 “Processing vertical size disparities in distinct depth planes” *Journal of Vision* 12(8):10 1-11

Duke P A, Oruc I, Qi H, Backus B T, 2006 “Depth aftereffects mediated by vertical disparities: evidence for vertical disparity driven calibration of extraretinal signals during stereopsis” *Vision Research* 46 228-241

Farell B, 1998 “Two-dimensional matches from one-dimensional stimulus components in human stereopsis” *Nature* 395 689-693

Fleet D J, Wagner H, Heeger D J, 1996 “Neural encoding of binocular disparity: energy models, position shifts and phase shifts” *Vision Research* 36 1839-1857

Gårding J, Porrill J, Mayhew J E, Frisby J P, 1995 “Stereopsis, vertical disparity and relief transformations” *Vision Research* 35 703-722

Gillam B, Chambers D, Lawergren B, 1988 “The role of vertical disparity in the scaling of stereoscopic depth perception: an empirical and theoretical study” *Perception & Psychophysics* 44 473-483

Greenwald H S, Knill D C, 2009 “Orientation disparity: a cue for 3D orientation?” *Neural Computation* 21 2581-2604

Harris J M, Chopin A, Zeiner K M, Hibbard P B, 2012 “Perception of relative depth interval: systematic biases in perceived depth” *The Quarterly Journal of Experimental Psychology* 65 73-91

Howard I P, Rogers B J, 1995 *Binocular Vision and Stereopsis* (New York: Oxford University Press)

Howard I P, Rogers B J, 2012 *Perceiving in Depth, Volume 2. Stereoscopic Vision* (New York: Oxford University Press)

Julesz B, 1971 *Foundations of Cyclopean Perception* (Chicago: University of Chicago Press)

Kaneko H, Howard I P, 1996 “Relative size disparities and the perception of surface slant” *Vision Research* 36 1919-1930

Kaneko H, Howard I P, 1997 “Spatial limitation of vertical-size disparity processing” *Vision Research* 37 2871-2878

Koenderink J J, van Doorn A J, 1976 “Geometry of binocular vision and a model for stereopsis” *Biological Cybernetics* 21 29-35

Matthews N, Meng X, Xu P, Qian N, 2003 “A physiological theory of depth perception from vertical disparity” *Vision Research* 43 85-99

Mitsudo H, 2007 “Illusory depth induced by binocular torsional misalignment. *Vision Research”* 47 1303-1314

Mitsudo H, 2012 “A minimal algorithm for computing the likelihood of binocular correspondence” *Japanese Psychological Research* 54 4-15

Mitsudo H, Kaneko H, Nishida S, 2009 “Perceived depth of curved lines in the presence of cyclovergence” *Vision Research* 49 348-361

Morgan M J, Castet E, 1997 “The aperture problem in stereopsis” *Vision Research* 37 2737-2744

Ogle K N, 1950 *Researches in Binocular Vision* (London: W. B. Saunders)

Porrill J, Frisby J P, Adams W J, Buckley D, 1999 “Robust and optimal use of information in stereo vision” *Nature* 397 63-66

Rogers B J, Bradshaw M F, 1993 “Vertical disparities, differential perspective and binocular stereopsis” *Nature* 361 253-255

Serrano-Pedraza I, Read J C A, 2009 “Stereo vision requires an explicit encoding of vertical disparity” *Journal of Vision* 9(4):3 1-13

van Ee R, Schor C M, 2000 “Unconstrained stereoscopic matching of lines” *Vision Research* 40 151-162

Vlaskamp B N, Filippini H R, Banks M S, 2009 “Image-size differences worsen stereopsis independent of eye position” *Journal of Vision* 9(2):17 1-13

Weiss Y, Simoncelli E P, Adelson E H, 2002 “Motion illusions as optimal percepts” *Nature Neuroscience* 5 598-604

Mitsudo, Sakai, & Kaneko Figure 1

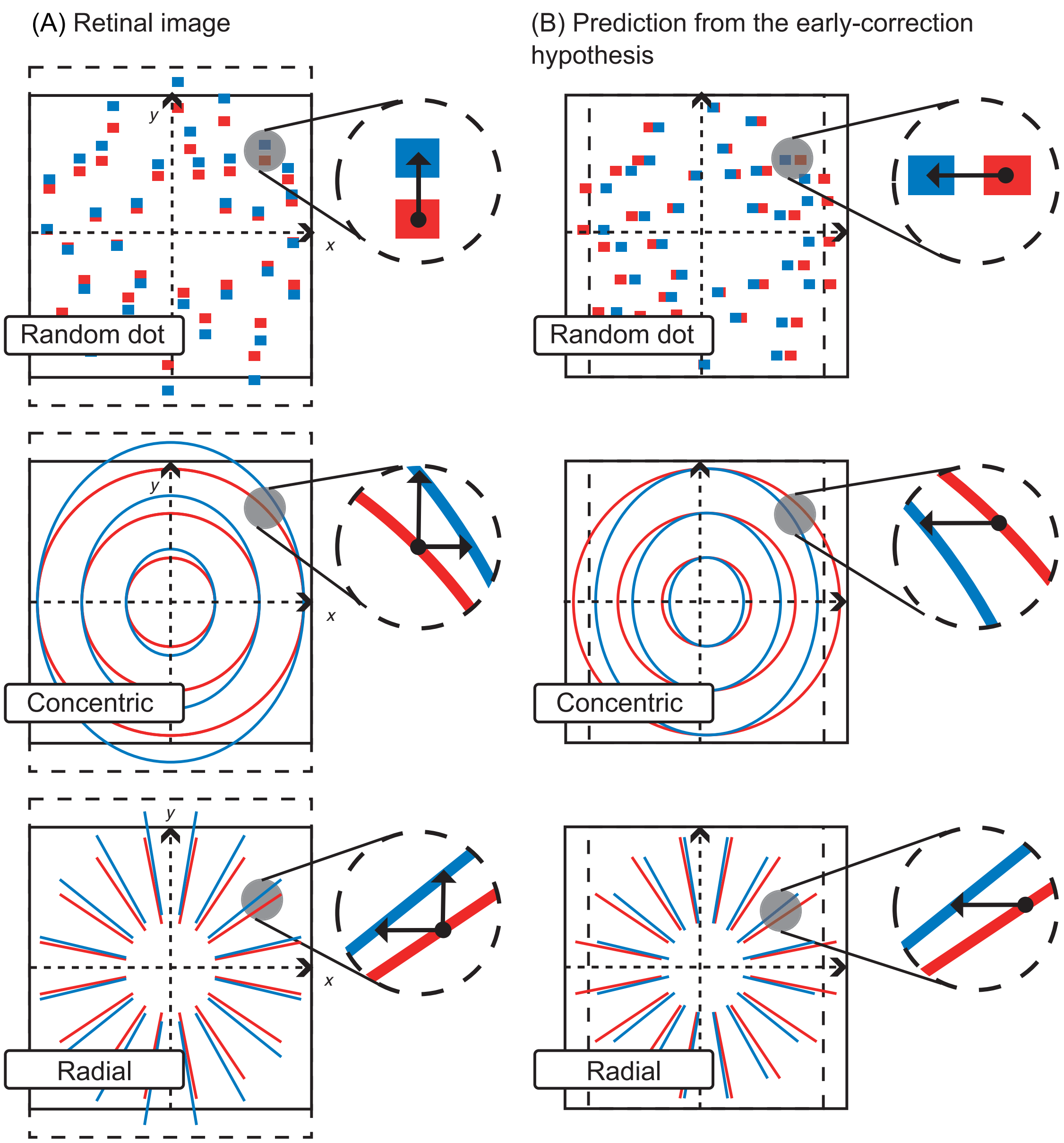

Figure 1. Schematic illustration of the random-dot, concentric, and radial patterns with a vertical size ratio of 105% (top, middle, and bottom rows, respectively). The right eye' s image is shown in red and outlined with a solid line, and the left eye' s is shown in blue and outlined with a dashed line. Arrows within circles represent binocular disparities. (A): actual retinal images. (B): corrected images predicted from the early-correction hypothesis. In (B), the left eye' s image is magnified by –4.76%. The stimuli and the size ratios are not drawn to scale.

Mitsudo, Sakai, & Kaneko Figure 2

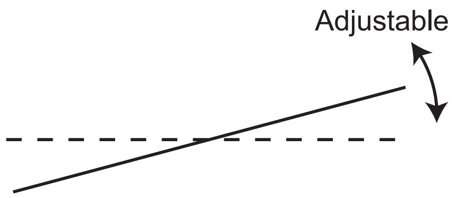

Figure 2. Schematic illustration of the matching stimulus. The arrow represents the adjustable line for reporting perceived slant.

Mitsudo, Sakai, & Kaneko Figure 3

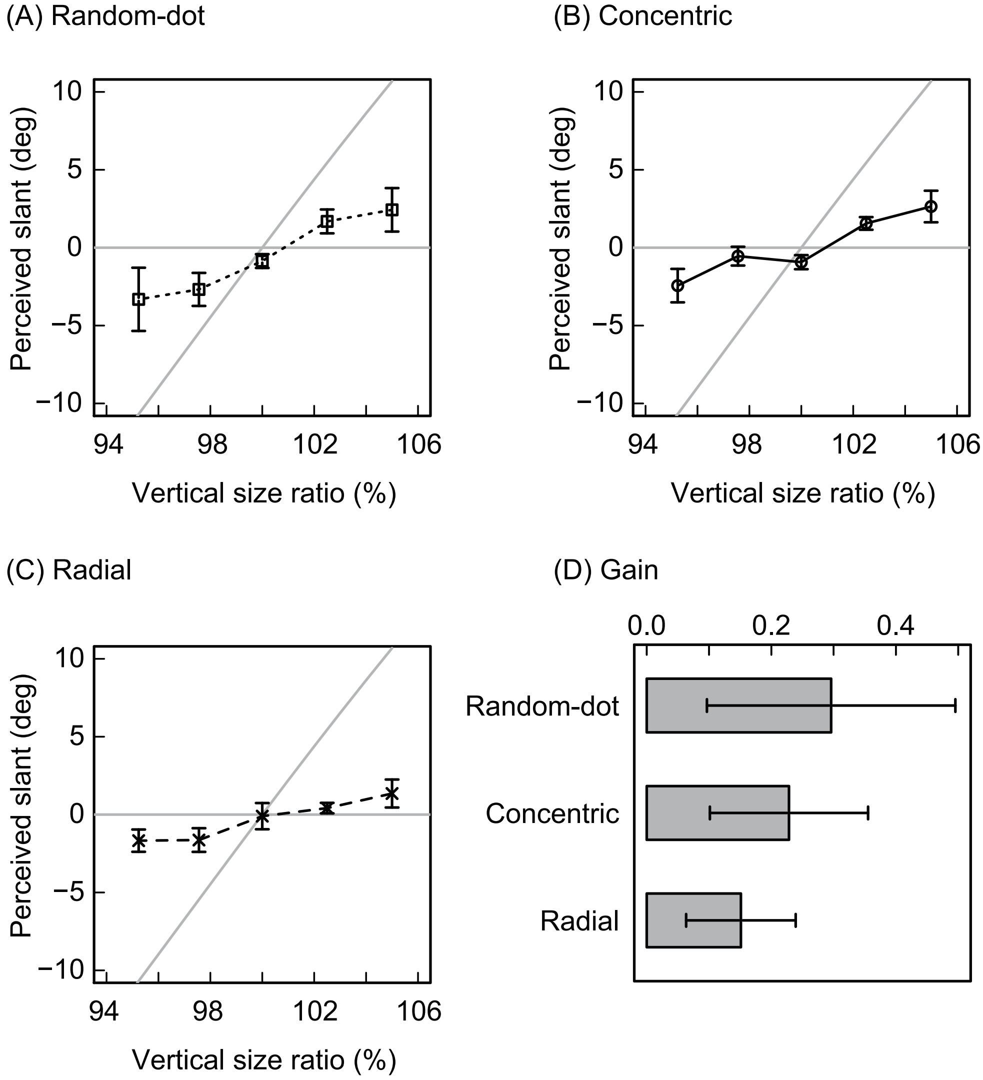

Figure 3. Results of Experiment 1. (A-C): Mean perceived slant as a function of vertical size ratio averaged over the ten observers. (D): Mean gains for the random-dot, concentric, and radial patterns. Error bars in (A-D) represent standard error.

Mitsudo, Sakai, & Kaneko Figure 4

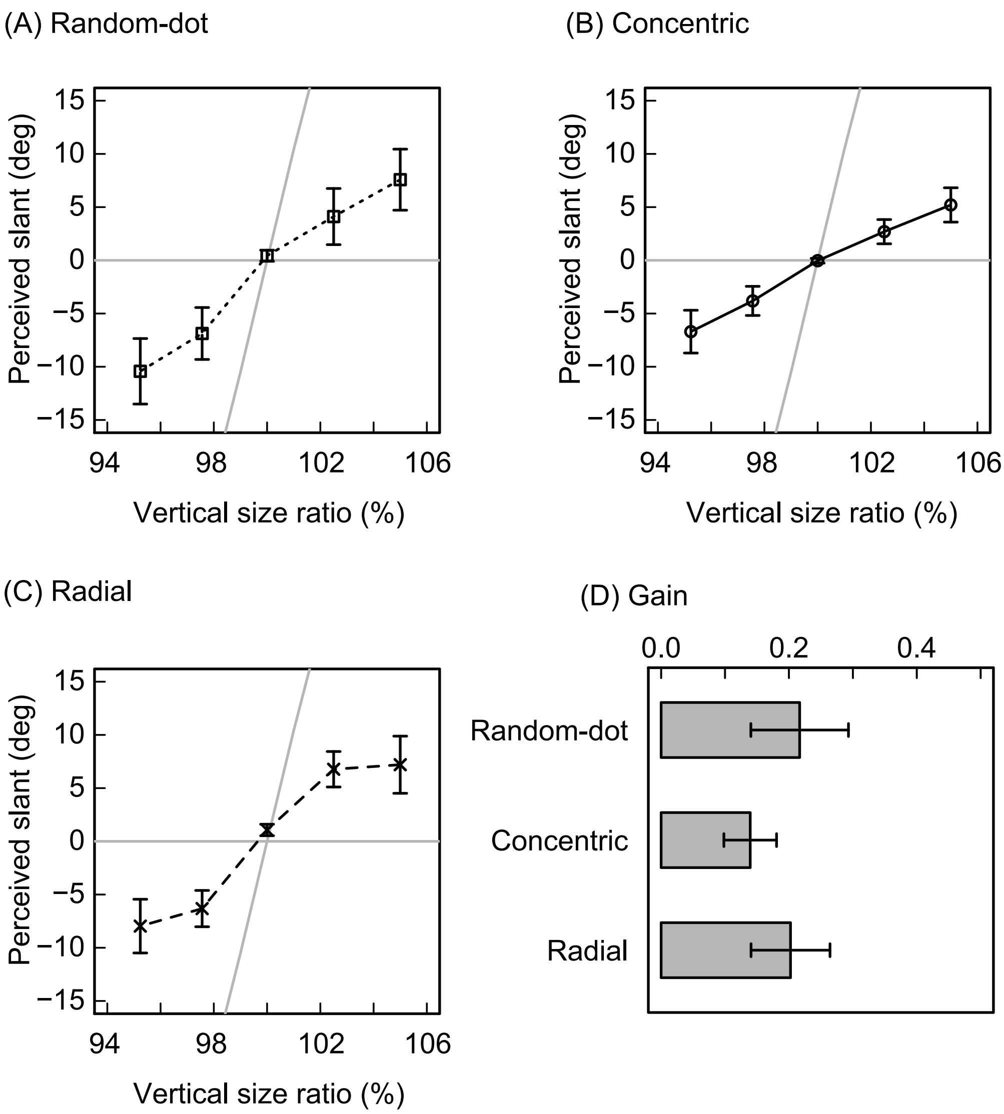

Figure 4. Results of Experiment 2. (A-C): Mean perceived slant as a function of vertical size ratio averaged over the five observers. (D): Mean gains for the random-dot, concentric, and radial patterns. Error bars in (A-D) represent standard error.

Mitsudo, Sakai, & Kaneko Figure 5

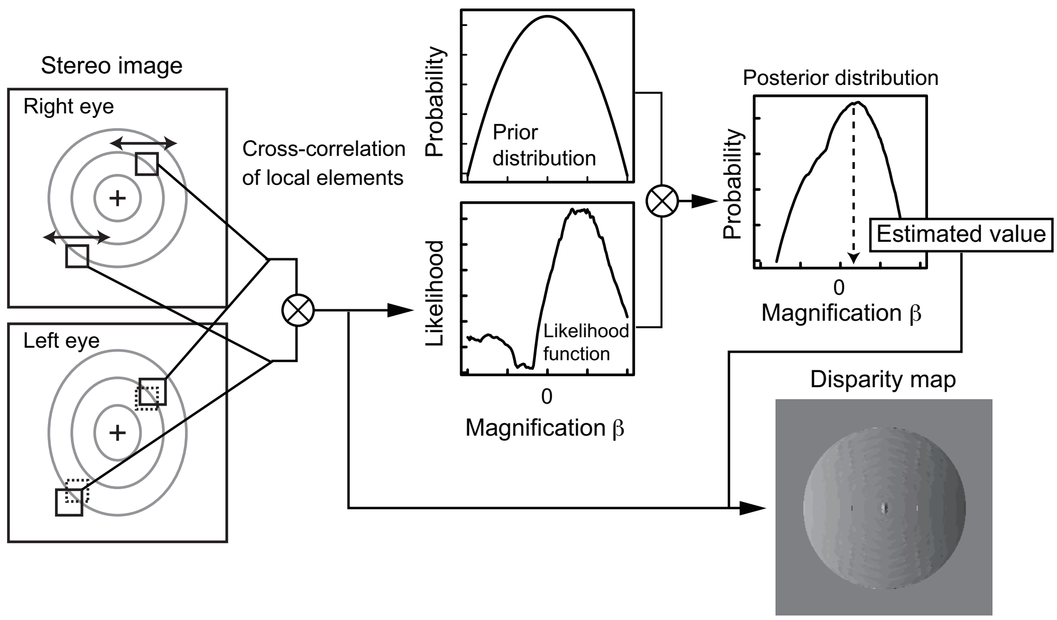

Figure 5. Schematic representation of the Bayesian model. The width (ie s.d.) of the prior distribution does not represent an actual value used in our simulations. See Section 4 for details.

Mitsudo, Sakai, & Kaneko Figure 6

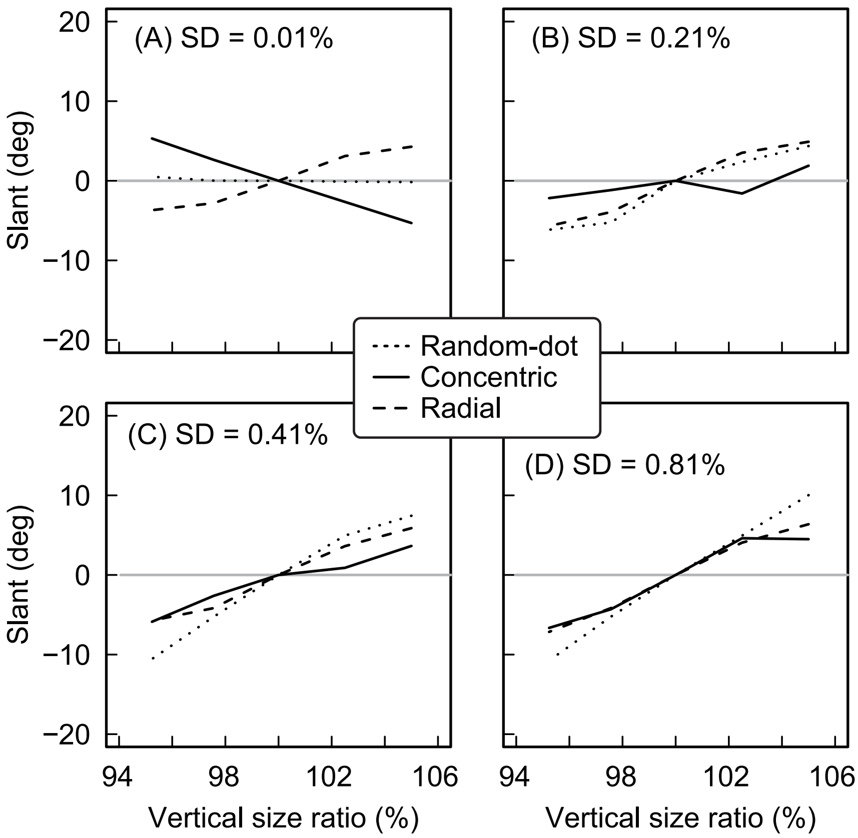

Figure 6. Simulation results of the Bayesian model. (A-D): Slants predicted with several values of the s.d. of the prior distribution of image magnification (0.01, 0.21, 0.41, and 0.81%, respectively).

Mitsudo, Sakai, & Kaneko Figure C

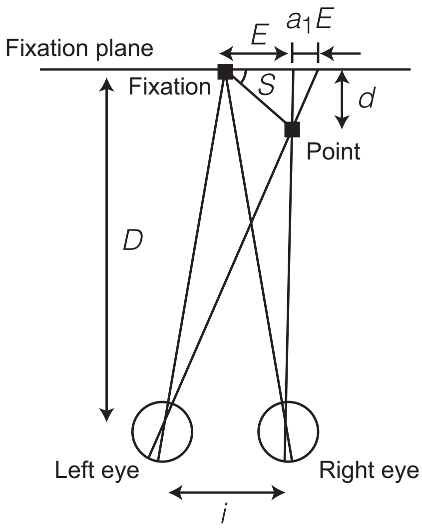

Figure C. A schematic illustration of the relation between slant and horizontal disparity in the model.