Author preprint, IEEE TVCG https://doi.org/10.1109/TVCG.2022.3203098

The impacts of lens and stereo camera separation on perceived slant in Virtual Reality head-mounted displays

Jonathan Tong*
Centre for Vision Research
York University

Robert S Allison†
Centre for Vision Research
York University

Laurie M Wilcox‡
Centre for Vision Research
York University

ABSTRACT

Stereoscopic AR and VR headsets have displays and lenses that are either fixed or adjustable to match a limited range of user inter-pupillary distances (IPDs). Projective geometry predicts a misperception of depth when either the displays or virtual cameras used to render images are misaligned with the eyes. However, misalignment between the eyes and lenses might also affect binocular convergence, which could further distort perceived depth. This possibility has been largely ignored in previous studies. Here, we evaluated this phenomenon in a VR headset in which the inter-lens and inter-axial camera separations are coupled and adjustable. In a baseline condition, both were matched to observers' IPDs. In two other conditions, the inter-lens and inter-axial camera separations were set to the maximum and minimum allowed by the headset. In each condition, observers were instructed to adjust a fold created by two intersecting, textured surfaces until it appeared to have an angle of 90°. The task was performed at three randomly interleaved viewing distances, monocularly and binocularly. In the monocular condition, observers underestimated the fold angle and there was no effect of viewing distance on their settings. In the binocular conditions, we found that when the lens and camera separation were less than the viewer's IPD, they exhibited compression of perceived slant relative to baseline. The reverse pattern was seen when the lens and camera separation were larger than the viewer's IPD. These results were well explained by a geometric model that considers shifts in convergence due to lens and display misalignment with the eyes, as well as the relative contribution of monocular cues.

Index Terms:

Human-centered computingVirtual reality; Human-centered computingEmpirical studies in HCI; Computing methodologiesPerception; Computing methodologiesVirtual reality

1 INTRODUCTION

With the growing popularity of stereoscopic-3D head-mounted displays (HMDs) for Augmented and Virtual Reality (AR and VR), there is a corresponding increase in interest in the the misperception of space due to unnatural viewing geometry in such devices [1, 2, 5, 9, 10]. One common issue arises when either the separation of the display optics or the virtual cameras used to render the scene do not match the user's inter-pupillary distance (IPD) [16, 23]. The IPD, which is the distance between the two eyes' pupils, relates binocular cues like convergence and disparity to depth [12, 29]. Binocular disparity is the difference between the retinal images due to two eyes' vantage points, while binocular convergence is the angle the eyes rotate to bring a point in depth into focus. Under natural viewing conditions, the visual system develops a mapping between these binocular cues and depth magnitude, according to its own IPD. In atypical viewing conditions, such as when viewing stereoscopic images that are rendered or captured with cameras at positions deviating from the viewer's eyes, the visual system will receive binocular cues that conflict with the 3D-scene layout. That is, the visual system will apply the mapping between binocular-cues and depth appropriate for natural viewing, without compensating for the mismatch in position between the cameras and the eyes. This, in turn, is predicted to result in a systematic and consistent misperception of depth, especially if feedback is unavailable [9, 10, 16, 19, 23]. Generally, in the field of stereoscopic imaging, the term "telestereopsis" is used to describe situations in which displayed content is captured by cameras with an inter-axial separation greater than the viewer's IPD [3, 4, 8, 13–15, 20, 24, 25]. This configuration magnifies perceived disparity, enhancing the sense of depth. In contrast, the term, "microstereopsis" is used to describe situations in which viewed content is captured by cameras with an inter-axial separation smaller than an observer's IPD [18, 27, 28]. This configuration decreases perceived disparity, minimizing the sense of depth. According to models based on projective geometry, perceived depth is expanded about the projection plane in telestereoscopic viewing, and compressed about the projection plane in microstereoscopic viewing. In the context of surface slant perception, telestereopsis is predicted to result in an overestimation of slant (biased away from frontoparallel) and microstereopsis is predicted to result in an underestimation of slant (biased towards frontoparallel). These predictions apply specifically to common viewing scenarios in which objects at or near the screen are rendered with zero disparity (the zero-parallax plane).

When the zero-parallax plane is placed at infinity, as in many VR HMDs with dual displays, the geometry predicts unbiased slant perception. Object size and depth are uniformly scaled such that slant does not change. Although projective geometry does not predict biases in perceived slant under this condition, it does predict a systematic bias in absolute distance perception. This is due to the fact that the lines of sight from the observer's left and right eyes (Fig. 1 solid red and blue lines) are constrained to follow paths parallel to the lines of sight of the left and right cameras to modeled object points in depth (Fig. 1, solid orange lines). The point at which the lines of sight for each eye intersect is where the observer perceives the object point (Fig. 1). Therefore, in telestereoscopic viewing with zero-parallax at infinity, the lines of sight from the observer's eyes converge at a point nearer in space than the modeled object point; the observer is predicted to underestimate object distance. Conversely, in microstereoscopic viewing with zero-parallax at infinity, the lines of sight from the observer's eyes converge at a point farther in space than the modeled object point; the observer is predicted to overestimate object distance.

Although the geometry described above makes specific predictions regarding binocular depth perception under conditions of telestereopsis and microstereopsis, it does not account for the possibility that binocular convergence might also be influenced by the placement of HMD optics (typically spherical magnifying lenses) and displays relative to the eyes. When viewing a virtual scene in an HMD with optics and displays that are coupled and misaligned from the viewer's eyes, the left and right eyes' lines of sight may deviate from parallel to the cameras' lines of sight, due to a combination of

*e-mail: tongj86@yorku.ca
†e-mail:allison@eecs.yorku.ca
‡e-mail:lwilcox@yorku.ca

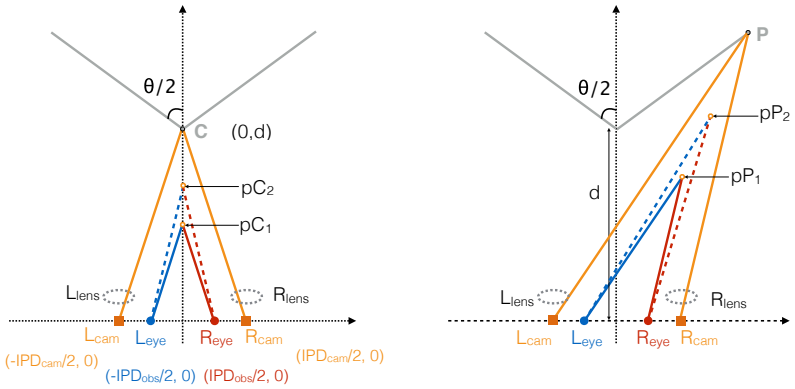

Figure 1: An illustration of the model's viewing geometry for a situation in which the HMD's IPD exceeds the observer's IPD; Left: the lines of sight (orange) from left and right-cameras to the fold-corner, C, and the lines of sight from the eyes, either following paths parallel to those of the cameras (solid red and blue lines) or deviating from these paths due to the influence of the position of the optics i.e. prismatic effect (dotted red and blue lines). The lines of sight for each eye intersect at the perceived position of the corner (p_{C1} and p_{C2}). Right: the same as in the figure on the left, only for a point on the surface, P. The lines of sight for each eye intersect at the perceived position of the surface point (p_{p1} and p_{p2})

the prismatic effect [26] and a shift in the center of the display (Fig 1, dashed red and blue lines): when the separation of the optics is less than the observer's IPD, the eyes must converge more to look at center of the displays; conversely, when the inter-axial separation of the optics is greater than the observer's IPD, the eyes must diverge to look through the center of the displays. Based on this geometry, we developed a model that proposes a dependence of binocular convergence on the degree of optical misalignment with the eyes by shifting convergence proportional to the difference between inter-axial lens separation and IPD (Fig. 1). The model predicts an underestimation of slant (perceived slant biased towards frontoparallel) when both inter-axial camera and lens separations are smaller than the observer's IPD; conversely, the model predicts that surface slant will be overestimated (perceived slant biased away from frontoparallel) when both inter-axial camera and lens separations are larger than an observer's IPD.

Here, we tested the predictions of our model by using a VR HMD with an adjustable "IPD". In this particular HMD, the adjustment of inter-axial camera and lens separations are coupled. We measured baseline performance, with both camera and lens separations matched to observers' IPDs, in a fold-angle adjustment task (Fig. 3) in which observers were asked to adjust the angle between two intersecting surfaces until they appeared to be perpendicular to one-another (an angle of 90°). The same task was carried out with the camera and lens separations set to the maximum and minimum allowed by the HMD, in turn producing mismatches with observers' IPDs. We compared task performance between baseline and "IPD mismatch" conditions and measured biases in perceived slant. Linear regressions revealed significant correlations between IPD mismatch and the adjusted fold angle relative to baseline. Our geometric model predicted the direction of the correlation between IPD mismatch and perceived bias in slant, however it did not fit the data as well as the linear regressions; the model predicted a much stronger effect especially at the farther viewing distances.

Our original model, as described above, only accounts for binocular cues to surface slant, such as convergence and disparity. However, monocular cues, such as foreshortening, and gradients of texture size and density, also play a substantial role in slant perception [11]. We reasoned that an increased reliance on monocular cues could mitigate the biases predicted by the perturbance of binocular cues due to IPD mismatch. To account for the contribution of monocular cues, we measured observers' performance on the fold-angle adjustment task with the stimulus only presented to the right-eye. In this condition, observers must rely on monocular cues alone to judge the slant of the surfaces, allowing us to estimate the contribution of monocular cues within a linear cue-combination framework. The cue-combination extension of our model produced better fits than the original model.

2 Methods

2.1 Participants

A total of 14 participants (mean age: 33 years; age range: 20-59 years) participated in the study. The 7 male participants had a mean IPD of 63 mm (SD = 4 mm) and the 7 female participants had a mean IPD of 62.6 mm (SD =2.9 mm). Before any experiments were run, we measured each participant's IPD using a digital PD (pupil distance) meter pupilometer (Newtry CP-32BT), to the nearest mm. All experimental protocols were approved by the York University Research Ethics Board.

2.2 Apparatus

We carried out experiments on the Oculus Quest VR headset (horizontal field of view: 90°; resolution: 1440 × 1600 per eye; refresh rate: 72 Hz). The experiment task was run as a stand-alone application developed in Unity 2019 and C-sharp. Observers used the right-hand Oculus Touch controller to register their responses. The Oculus Quest 1 headset allowed for adjustment of IPD within a range of 59 mm to 71 mm, in 1 mm steps, by movement of a slider. In a coupled manner, the IPD slider adjusted the center-to-center spacing between the lenses and the inter-axial virtual camera separation (automatically applied by the HMD driver) to the specified value.

2.3 Procedure

The headset "IPD" setting was adjusted to a predetermined value at the beginning of each testing block (experimental condition). In the baseline condition, the inter-lens and inter-axial camera separations were matched to the observers' IPDs to the nearest mm (matched-IPD setting). In two separate IPD-mismatched conditions, the inter-lens and inter-axial camera separations were set to 59 mm (low IPD-setting) and 71 mm (high-IPD setting). For observers who had extreme IPDs of 59 mm (n = 2) or an IPD of 71 mm (n = 1), their mismatch conditions consisted of setting the headset IPD to 65 mm and to the other extreme IPD setting (71 mm for observers with IPD = 59 mm; 59 mm for observer with IPD = 71 mm).

In a monocular condition, the headset-IPD was matched to the observers. However, the virtual scene was only displayed to the

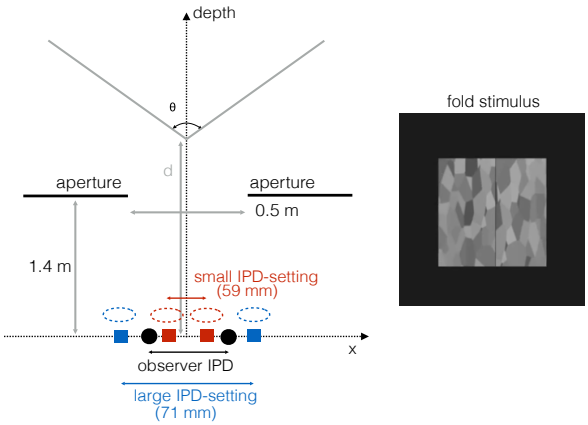

Figure 2: Left: The viewing geometry in the fold-angle adjustment task; the corner of the fold was a distance, d (1.5, 2, or 2.5 m) away from the observer. The fold angle appeared behind an aperture, 0.5 m in width, 1.4 m away. The lenses in the headset (dotted ellipses), the microdisplays (not shown) and virtual cameras (squares) were coupled. Observers adjusted the fold angle, θ, until it appeared to be 90°. Right: Screenshot, taken from the right-eye view, showing the voronoi-textured fold stimulus

right-eye. A uniform grey background was displayed to the left-eye. Observers completed the monocular condition block either before or after the binocular blocks, with the order of the monocular block counterbalanced. The testing order of the binocular conditions was counterbalanced for all observers with non-extreme IPDs (59 mm <IPD <71 mm). The remaining observers with extreme IPDs (59 mm or 71 mm) ran the blocks in randomized order.

Each experimental block consisted of a Fold Adjustment task. The virtual scene (Fig. 2) appeared behind a square aperture, 20.25° in width (0.5m wide, and tall, at a distance of 1.4 m). In the task, observers were first instructed to fixate on a central fixation cross that briefly appeared, for 1 second, at a viewing distance of either 1.5 m, 2 m or 2.5 m away. After the fixation cross disappeared, the fold stimulus appeared at the same distance as the fixation-cross, with the corner of the fold at either 1.5 m, 2 m or 2.5 m away (Fig. 2, right). The stimulus consisted of two intersecting surfaces, textured with a voronoi pattern (with a tessellation frequency of approximately 0.5 cycles/degree); the size of the fold , and associated texture, were scaled according to viewing distance so as to maintain a consistent image size. The initial angle between the two surfaces was randomly set at a value between 40° and 140°. Observers were instructed to adjust the fold-angle by toggling the joystick (Oculus Touch controller) left and right to make the angle smaller or larger, until the angle appeared to be 90° (each panel apparently 45° away from frontoparallel). Once the observers were confident they adjusted the angle to 90° they pressed either the A or B button on the controller to register their response and move on to the next trial. 20 adjustments were made for each of the 3 viewing distances, for a total of 60 trials per block.

2.4 Analysis

We computed the mean adjusted angle for each observer in each condition: baseline, low IPD-setting, high IPD-setting, monocular and the three viewing distances (1.5 m, 2 m and 2.5 m) within each condition. Difference scores were calculated between the mean adjusted angle for the base-line condition and each of the IPD-mismatch conditions: baseline - low IPD-setting, and baseline - high IPD-setting. For each viewing distance, we ran a linear regression analysis with the difference scores as the dependent variable and the IPD-mismatch (participant IPD – HMD IPD) as the independent variable. A significant correlation would suggest an effect of IPD-mismatch on the difference of perceived fold-angle relative to baseline.

We carried out the linear-regression and subsequent model-fitting analyses with outlier data-points omitted. Outliers were classified as any data point in a particular viewing distance condition whose Cook's Distance [7] was greater than 3 times the mean Cook's Distance. Overall, three outlier data-points were omitted from each of the distance conditions.

The linear-regressions and outlier analysis was carried out in R (Version 0.99.893) using the standard linear models function. The geometric model simulations and fitting were run in Matlab (R2015b).

2.5 Computational model

We fit the regression data with a computational model as described below. Refer to fig. 1 for a visualization of the modeled viewing geometry and notation used throughout this section. For simplicity, we reduced the geometry to two-dimensions, with x-coordinates representing horizontal position and y-coordinates representing depth. The origin of the coordinate system was defined to be exactly halfway between the observer's left and right eyes, separated by IPD_{obs},with the coordinates:

$$(x_{Leye}, y_{Leye}) = \left(\frac{-IPD_{obs}}{2}, 0\right)$$
$$(x_{Reye}, y_{Reye}) = \left(\frac{IPD_{obs}}{2}, 0\right),$$

respectively. The left and right camera positions had coordinates defined by the inter-axial camera separation, and which were:

$$(x_{Lcam}, y_{Lcam}) = \left(\frac{-IPD_{cam}}{2}, 0\right)$$
$$(x_{Rcam}, y_{Rcam}) = \left(\frac{IPD_{cam}}{2}, 0\right),$$

respectively. The corner of the fold, C, where the left and right surfaces intersected, was always aligned with the midline between the eyes and in front of the observer, at a distance, d. Therefore, the coordinates of the corner for the near, mid and far viewing distances were $(x_C, y_C) = (0, 1.5), (x_C, y_C) = (0, 2)$, and $(x_C, y_C) = (0, 2.5)$, respectively. The furthest visible point, P, on the right surface, given a viewing angle of ω = 20.25° (the width of the aperture) was considered for analysis; the exact coordinates of this point depended on the angle between left and right surfaces, θ. The coordinates of this point were computed as:

$$(x_P, y_P) = \left(\frac{y_c}{\frac{1}{tan(\omega/2)} - \frac{1}{tan(\theta/2)}}, \frac{\frac{y_c}{tan(\omega/2)}}{\frac{1}{tan(\omega/2)} - \frac{1}{tan(\theta/2)}}\right) \quad (1)$$

Since the fold angle was symmetric, as were the position of points on left and right surface, we need only consider points on one side.

Given the coordinates of the eyes, the cameras, fold-corner and surface-point, described in the previous paragraph, the linear equations ($y = mx + b$) describing the lines of sight from each camera to the object-points (fold-corner and surface-point) were determined. See table 1 for a summary of the slopes (m) and y-intercepts (b) of the linear equations describing the lines of sight from camera to object points.

To model the influence of coupled lens and display offset from the eyes, the lines of sight for each eye depended on the difference between the observer's IPD and the inter-axial separation of the

Table 1: The linear equation parameters describing each camera's line of sight for each object point

	Left Camera	Right Camera
Corner	$m_{LcamC} = \frac{-y_C}{x_{LCam}}$	$m_{RcamC} = \frac{-y_C}{x_{RCam}}$
	$b_{LcamC} = y_C$	$b_{RcamC} = y_C$
Point	$m_{LcamP} = \frac{y_P}{(x_P - x_{LCam})}$	$m_{RcamP} = \frac{y_P}{(x_P - x_{Rcam})}$
	$b_{LcamP} = -m_{LcamP} * x_{Lcam}$	$b_{RcamP} = -m_{RcamP} * x_{RCam}$

Table 2: The linear equation parameters describing each eye's line of sight for each object point

	Left Eye	Right Eye
Corner	$m_{LeyeC} = \frac{1}{tan(tan^{-1}(1/m_{LcamC}) + \gamma)}$	$m_{ReyeC} = \frac{1}{tan(tan^{-1}(1/m_{RcamC}) - \gamma)}$
	$b_{LeyeC} = \frac{IPD_{obs}}{2} m_{LeyeC}$	$b_{ReyeC} = -\frac{IPD_{obs}}{2} m_{ReyeC}$
Point	$m_{LeyeP} = \frac{1}{tan(tan^{-1}(1/m_{LcamP}) + \gamma)}$	$m_{ReyeP} = \frac{1}{tan(tan^{-1}(1/m_{RcamP}) - \gamma)}$
	$b_{LeyeP} = -m_{LeyeP} * x_{Leye}$	$b_{ReyeP} = -m_{ReyeP} * x_{Reye}$

lenses (equivalent to the inter-axial camera separation, $IPD_{lens} = IPD_{cam}$). Specifically, the offsets introduces a lateral shift, both through the prismatic effect and displacement of the center of the displays relative to the eyes, that the eyes must compensate for by adjusting their vergence angle. The lateral shift , Δ, is computed using the principle of similar triangles (fig. 3), taking the lens focal length, f, and screen distance from the lens, o, into consideration:

$$\Delta = \frac{IPD_{obs} - IPD_{lens}}{2} \left(\frac{f - o}{f} \right) \tag{2}$$

While we do not have details of the optical design of the Quest we used the focal length of the Oculus DK2 lens ($f = 0.045$ m), the design focal distance of 1.3 m, and the approximate magnification of the image to estimate $o = 0.0435$ m, numbers that are consistent with the general physical dimensions of the Quest.

This lateral shift Δ was expressed in terms of visual angle:

$$\gamma = tan^{-1} \left(\frac{\Delta}{o} \right) \tag{3}$$

The visual angle of the lateral offset, γ, was added to the angle complementary to the slope angle of the left eye's line of sight and subtracted from the angle complementary to the slope angle of the right eye's line of sight. Table 2 summarizes the computations for each eye's line of sight (slope and y-intercept) for each object point (surface point and corner), given the change in vergence angle, γ.

For the fold-corner, the left and right eye's lines of sight will intersect at the same y-intercept value; this point of intersection is where the corner of the fold should be perceived,

$$(x_{pC}, y_{pC}) = \frac{IPD_{obs}}{2} * m_{LeyeC} = -\frac{IPD_{obs}}{2} * m_{ReyeC} \tag{4}$$

Once we determined the linear equations describing the lines of sight of each eye for the surface-point, we solved for the coordinates of the point where these lines intersect (where the surface-point is perceived). For the linear equations describing the lines of sight, written in the following form:

$$m_{LeyeP} x + y + b_{LeyeP} = 0$$
$$m_{ReyeP} x + y + b_{ReyeP} = 0,$$

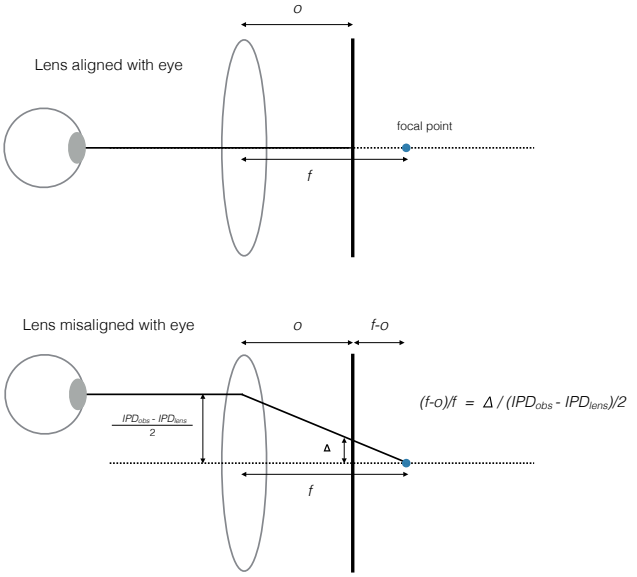

Figure 3: Top: When the lens is aligned with the eye, no vergence is needed to look through the center of the lens to the display center. Bottom: When the lens is misaligned with the eye, the vergence angle of each eye changes to compensate for a lateral shift Δ, between the display center (equivalent to optical axis) and where the refracted line of sight, due to the prismatic effect, intersects the display. The principle of similar triangles can be used to solve for Δ, given the expression on the right.

Cramer's rule can be used to obtain the solution for the point, (x, y), shared by both of the linear equations:

$$(x, y) = \left(\frac{b_{LeyeP} - b_{ReyeP}}{m_{ReyeP} - m_{LeyeP}}, \frac{b_{LeyeP} * m_{ReyeP} - b_{ReyeP} * m_{LeyeP}}{m_{ReyeP} - m_{LeyeP}} \right) \tag{5}$$

Eqn. (5) was used to solve for the position coordinates of the perceived surface-point (x_{pP}, y_{pP}). These coordinates, along with the coordinates of the perceived fold corner (x_{pC}, y_{pC}) were then used to compute the perceived fold-angle, θ_p:

$$\theta_p = 2tan^{-1} \left(\frac{x_{pP} - x_{pC}}{y_{pP} - y_{pC}} \right) \tag{6}$$

We ran model simulations of the experiment for comparison with the behavioural data by carrying out the following steps (Fig. 4, see numbered steps):

1. we used each observer's IPD for the model IPD_{obs}, and simulated the mismatch conditions by setting the model IPD_{cam} to the headset IPD in the small (59 mm) and large (71 mm) IPD settings

2. we used the tested viewing distances of d = 1.5 m, 2.0 m and 2.5 m as input in the model for the y-coordinate of the fold-corner y_C

3. for a range of input angles, θ = 20° to 180° in 0.1° steps, the coordinates of both the corner and surface points were used in the model to predict the perceived coordinates of the corner (x_{pC}, y_{pC}) and surface points (x_{pP}, y_{pP}) (Eqn. 5), which in turn were used to compute the predicted perceived angle, θ_p (Eqn. 6)

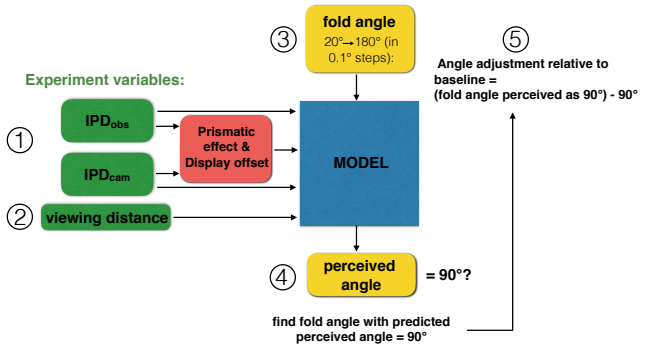

Figure 4: The procedure for model simulation and fitting, with numbered steps

4. the input angle, θ, that resulted in the perceived angle, θ_p, nearest 90° was taken as the model's predicted "adjusted angle" ($\theta_{predicted}$), which, in turn, was subtracted by 90° to compute the predicted deviation in adjusted angle relative to baseline (where $IPD_{cam} = IPD_{obs}$ was predicted to result in perfect performance).

5. The model's predicted deviation in adjusted angle relative to baseline was compared to that of the observer's for each condition (Fig. 6).

We computed the root mean squared error (RMSE) for the model prediction compared to the data, for each viewing distance (with outliers removed; see 2.4 Analysis).

3 Modeling the Contribution of Monocular Cues

The model's predicted adjustment angle, as described in the previous section, is solely based on binocular geometry. We extended the model to consider the contribution of monocular cues by using a linear cue-weighting approach [11]. We computed a weighted average of the basic model's output ($\theta_{predicted}$, fold angle that resulted in a perceived angle of 90°) and the overall mean adjusted angle experimentally observed in the monocular condition (θ_{mono}), for the appropriate viewing distance (Fig. 5, right). The weighted average was computed as follows:

$$\theta_{combined} = \theta_{predicted} w_{bino} + \theta_{mono} w_{bino} \quad (7)$$

The weights for the model's binocular estimate, w_{bino} , and the monocular adjustment angle, w_{mono} , could take on values between 0 and 1 and were constrained so that: $1 = w_{bino} + w_{mono}$. This constraint allowed us to vary only one of the weights; we chose to vary w_{mono} as a free-parameter for the cue combination analysis. w_{mono} was allowed to independently vary between viewing distance conditions, since there is evidence in the literature that the relative weighting of monocular cues changes with viewing distance [11].

In the binocular model described earlier, the predicted adjusted angle for the baseline condition was 90°; since in the baseline IPD_{obs} was equal to IPD_{cam}, there was no predicted bias in the baseline. However, in the extended cue-combination model, we had to consider the bias present in the monocular condition when predicting the baseline adjusted angle (monocular cues could also contribute in the baseline condition); the predicted baseline adjustment angle was computed by evaluating eqn. 7, above, with $\theta_{predicted} = 90°$.

The cue-combined estimates for the mismatch conditions were subtracted from the predicted baseline adjustment angle to give a predicted adjustment angle relative to baseline. In turn, this value, for each simulated condition, was compared to the measured adjusted angle relative to baseline in the corresponding condition, as described in the previous model-fitting section. We minimized the RMSE, and report the W_{mono} that resulted in the overall minimum RMSE.

4 Results

4.1 Binocular Experiment Results

Generally, in the fold-angle adjustment task, observers tended to underestimate the degree of slant in the surfaces, such that they consistently adjusted the fold angle less than 90° (each surface slanted more than 45° from frontoparallel). Furthermore, in binocular experiments, this tendency to underestimate the degree of slant became stronger as viewing distance was increased. This was true even when the headset IPD setting was matched to the observer's own IPD in the baseline condition (Fig. 5, left and middle). However, we were primarily interested in how observers perceived slant when the headset IPD did not match their own IPD, relative to when it did. Therefore, we computed the difference between their mean adjusted-angle in the baseline (matched-IPD) condition and their mean adjusted angle in each of the mismatched-IPD conditions: baseline adjustment – mismatched IPD adjustment. When the headset IPD-setting was greater than the observer's IPD, they generally overestimated the amount of slant and therefore adjusted the fold corner wider than in the baseline condition. Conversely, when the headset IPD-setting was less than the observer's IPD, they generally underestimated the amount of slant and therefore adjusted the fold corner narrower than in the baseline condition.

Linear regressions were performed for each of the three viewing distances, with the IPD-mismatch (observer IPD – headset IPD) as the independent variable and the deviation of angle adjustment relative to the baseline condition (baseline adjustment – mismatched IPD adjustment) as the dependent variable (Fig. 6). The linear regressions revealed significant positive correlations between IPD mismatch and the deviation of angle adjustment relative to the baseline condition, for the 1.5 m viewing distance (adjusted R^2= 0.40, F(1,20)= 14.87, $p < 0.01$; slope = 1.03, y-intercept = 5.75), the 2.0 m viewing distance (adjusted R^2= 0.30, F(1,20)= 10.16, $p < 0.01$; slope = 0.38, y-intercept = 1.83) and the 2.5 m viewing distance (adjusted R^2= 0.42, F(1,20)= 27.82, $p < 0.001$; slope = 0.65, y-intercept = 0.82).

4.2 Basic Model Predictions

The data in the regression analysis were further compared to our geometric model predictions (Fig. 6, red plots). For the near viewing distance of 1.5 m the model had an RMSE = 10.46°; for the mid viewing distance of 2.0 m the model had an RMSE = 11.19°; and for the far viewing distance of 2.5 m, the model had an RMSE = 12.84°. Overall, our basic model did not perform as well as the linear regression at predicting the effect of IPD mismatch on the adjustment angle relative to baseline. The model overestimates the effect of IPD mismatch on the adjusted fold angle relative to baseline, especially at the 2 m and 2.5 m viewing distances. Next, we considered the possibility that monocular cues might contribute to the task, and serve to mitigate the effect of IPD mismatch predicted by the model.

4.3 Monocular experiment results

Given that our basic model only generates predictions based on binocular cues, we were interested in investigating the contribution of monocular cues in slant perception, and the possibility that relying on monocular cues might mitigate the effects of IPD mismatch. Therefore, we had observers complete the fold angle adjustment task under monocular viewing conditions to get a baseline measurement of performance when no binocular cues are available. In the monocular condition, the mean adjustment angle (angle perceived as 90°) did not change across viewing distances (Fig. 5, right), unlike in binocular viewing conditions where the adjustment angle tended to decrease with increasing viewing distance. In the monocular

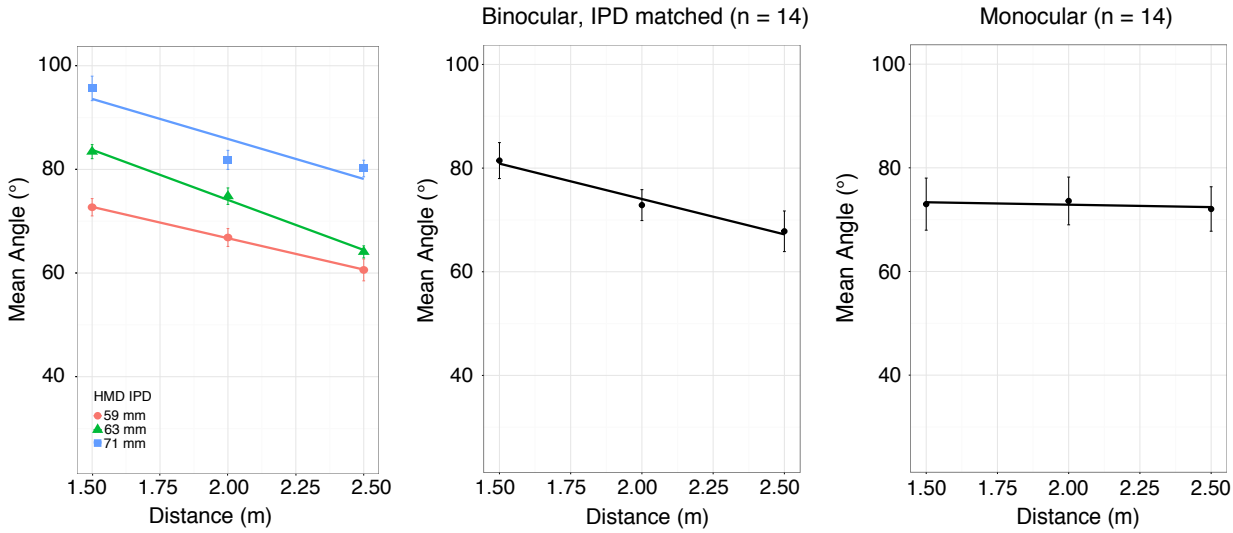

Figure 5: Left: the mean angle adjustments±SE (angle perceived as 90°) plotted as a function of the viewing distance for the matched IPD condition (green), large-IPD (71mm) condition (blue), and small-IPD (59 mm) condition (red) for a representative observer. The solid lines are the linear-fits for each condition. Middle: the mean angle adjustment±SE in the baseline (matched IPD) condition across all observers. The solid lines are the linear-fits for each condition. Right: the mean angle adjustment±SE in the monocular condition across all observers. he solid lines are the linear-fits for each condition.

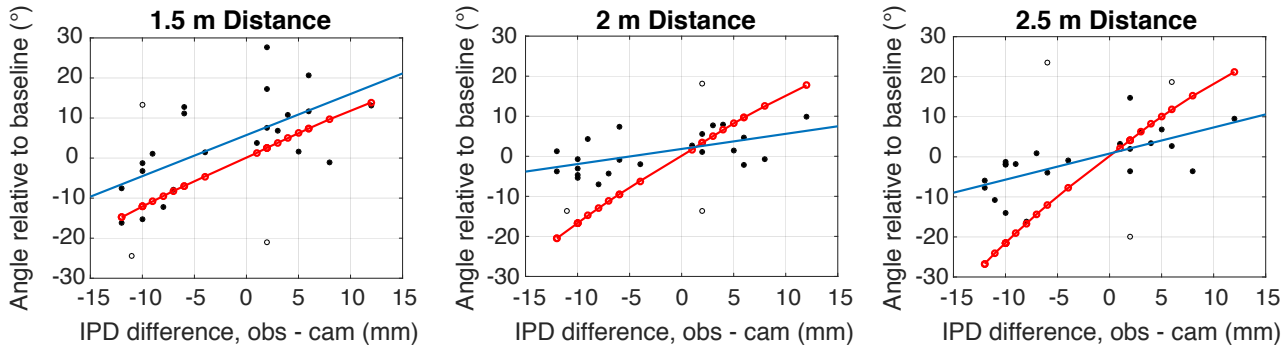

Figure 6: The deviation of angle adjustment relative to the baseline condition plotted as a function of IPD-mismatch, with the linear regression (blue line) and model (red line) predictions across the three viewing distance conditions. Open data points: outliers omitted from the regression and model-fitting analyses.

condition, just as in the binocular conditions, observers consistently adjusted the fold less than 90°.

4.4 Cue-combination Model fits

To model the contribution of monocular cues in the fold-angle adjustment task, we repeated the model-fitting procedure as previously described, within a linear cue combination framework. In this version of the model, the best-fit monocular weight for the 1.5 m viewing distance was $W_{mono} = 0.30$ (RMSE = 10.10°); the best-fit monocular weight for the 2 m viewing distance was $W_{mono} = 0.82$ (RMSE = 4.25°); and the best-fit monocular weight for the 2.5 m viewing distance was $W_{mono} = 0.71$ (RMSE = 5.36°). Thus, as predicted, the inclusion of monocular cues in a linear cue-combination framework improved model fits, especially for the 2 m and 2.5 m viewing distances (Fig. 7). In fact, the extended model predicted the results at a level comparable to the linear regression for the 2.5 m viewing distance condition

5 Discussion

Our psychophysical results show that a mismatch between an observer's IPD and the inter-axial separation of both the lenses and virtual cameras in an HMD results in a systematic misperception of slant. When the inter-axial separation of the lenses and virtual camera's exceeded an observer's IPD, surfaces appeared more slanted relative to when these parameters were matched to the user's IPD. Conversely, when the inter-axial separation of lenses and virtual camera's was less than the observer's IPD, surfaces appeared less slanted (less apparent depth) relative to when these parameters were matched to the user's IPD.

Previous geometric models predict a compression or expansion of perceived slant due to a mismatch between virtual camera separation and IPD, but only when the zero-parallax plane is some finite value away [9, 10]. Our model proposes an effect of lens misalignment on the baseline degree of vergence, and predicts the same pattern of biases in perceived slant we observed experimentally. Specifically, when the lenses (and displays) are separated by an amount greater than the user's IPD, their eyes must diverge slightly to look through

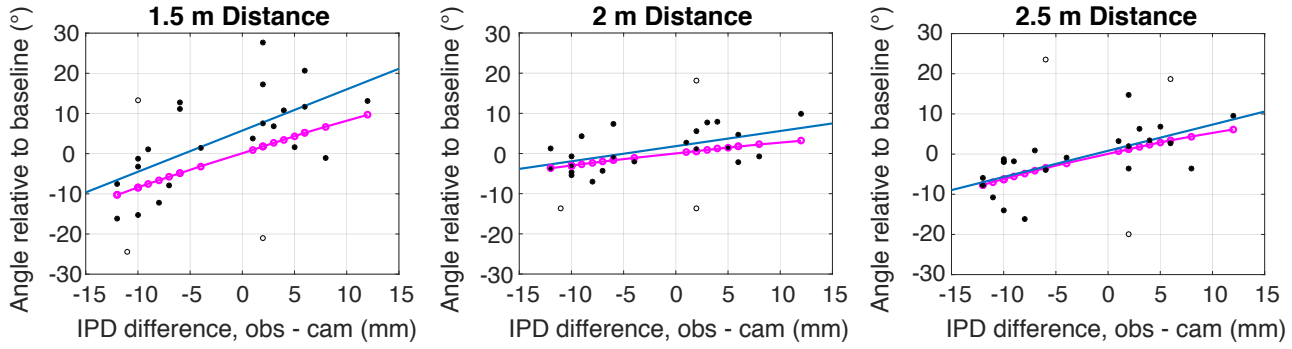

Figure 7: The deviation of angle adjustment relative to the baseline condition plotted as a function of IPD-mismatch, with the cue-combination model best-fits (magenta lines) and linear regression best-fits (blue lines) to the data. Open data points: outliers omitted from the regression and model-fitting analyses.

the lens at the display center. This slight divergence interacts with the convergence required to view corresponding points, as determined by the separation between cameras, and ultimately predicts an overestimation of surface slant. Conversely, when the lenses (and displays) are separated by an amount less than the user's IPD, their eyes must converge slightly to look through the lens at the display center. This slight convergence interacts with the convergence required to view corresponding points, as determined by the separation between cameras, and ultimately predicts an underestimation of surface slant.

Telestereoscopic and microscopic viewing conditions are known to introduce biases in perceived absolute distance [17], relative distance [2, 5], and slant [20]. Generally, depth and slant are underestimated in microstereoscopic viewing and overestimated in telestereoscopic viewing, in accordance with geometric model predictions [9, 10]. The experiments in these studies, however, involved displaying stimuli on a stereoscopic projector with a fixed screen distance, only meters away from the observer. In many VR headsets, however, the zero-parallax plane can often be placed at infinity. Further, given that most HMDs use magnifying lenses, it is also important to take their effects on vergence into consideration. As far as we are aware, this is the first study to systematically manipulate both virtual camera and lens inter-axial separations relative to the eyes in a VR headset to probe their affects on perceived depth. We should also note that the aforementioned studies tested a range of camera separations, with minimum and maximum values well beyond physiologically realistic range of IPDs. Our study focused exclusively on camera (and lens) separations within a typical physiological range of IPDs (59-71mm; [10]). Despite this, we were able to measure significant effects of mismatches on depth perception (specifically in perceived slant) within this range. These results suggest that even slight mismatches of the lenses, displays, or cameras to the eyes in stereoscopic HMDs may lead to measurable systematic misperception of depth.

As viewing distance increases, the relative reliability of binocular cues tends to diminish and therefore the reliance on monocular cues increases [11]. We modeled the contribution of monocular cues to the perception of slant in a fold-angle adjustment task by taking a weighted average of the model's predicted fold-angle perceived as 90°, based on binocular geometry, and the measured adjustment angle in the monocular condition. We allowed the monocular weight to vary between viewing distances, since it is also known that the relative reliability of monocular cues varies with viewing distance. Within the range of viewing distances we tested (1.5-2.5m) our linear cue-combination analysis revealed an increase in the best-fitting monocular weight between the 1.5 m to 2.0 m viewing distances, from a value of 30% to 82%. However, the best-fitting monocular weight fell from 82% to 71% between the 2.0 m to 2.5 m viewing distances. Though the increase in the weighting of monocular cues was not monotonic, it's clear that inclusion of this information substantially improved the model's ability to capture human performance. Further improvements would likely be seen as the model is developed to incorporate a wider set of monocular cues to depth/slant.

The misalignment between the lenses and the eyes may have other effects besides possibly modulating the amount of convergence needed to fixate near points. For example, the magnifying lenses typically introduce spatial distortions that are corrected in HMDs by applying an inverse pre-distortion. This pattern of spatial distortion changes when the eye is not located at the center of the lens. The resulting phenomenon, known as 'pupil-swim', has been demonstrated even when corrective pre-distortion is applied [6]. We cannot rule out that such geometric lens distortion may have played a role in the misperception of surface slant as the lenses became more misaligned with the eyes. In fact, in previous studies we have demonstrated that geometric lens distortion introduces biases in slant perception, especially in the periphery [21, 22]. The barrel distortion introduced by the magnifying lenses in most HMDs generally gives the appearance of concave curvature in planar surfaces: a frontoparallel surface would appear to slant towards the observer in the periphery. For surfaces that slant away towards the periphery, as in our fold angle stimulus, under-corrected barrel distortion could cause slant to be underestimated [21]. This might partially explain the baseline underestimation of slant we measured in both the monocular and IPD-matched conditions in the current study; it might also explain why the slant bias due to IPD mismatch appears to be asymmetric in the data, where the underestimation of slant is stronger than the overestimation of slant when IPD is mismatched. This asymmetry was reflected in linear regression best-fits with y-intercepts consistently above zero, which the geometric model could not account for since it predicts no bias with zero IPD mismatch.

6 Conclusion

In conclusion, we provide evidence that mismatch of the inter-axial camera and lens separations relative to a user's IPD, in a VR HMD, will result in systematic misperception of surface slant. The strength of these perceptual biases is proportional to the magnitude of mismatch between these head-set parameters and user IPD. These results were well-explained by a geometric model that, in addition to the scaling of binocular disparity, incorporates convergence changes due to the mismatch between headset parameters and users' IPD, and the contribution of monocular information. As noted above, we tested a relatively modest range of IPD mismatches, consistent with typical viewing scenarios; even so we see significant impact on perceived slant in depth. Ideally, headsets should incorporate adjustable lens IPD (as is the case in some models). In any case, developers of

AR and VR applications should consider these perceptual biases in situations where these parameters are fixed or when the adjustment of these parameters is limited.

REFERENCES

[1] R. S. Allison. The camera convergence problem revisited. In *Stereoscopic Displays and Virtual Reality Systems XI*, vol. 5291, pp. 167–178. SPIE, 2004.

[2] R. S. Allison and L. M. Wilcox. Perceptual tolerance to stereoscopic 3d image distortion. *ACM Transactions on Applied Perception (TAP)*, 12(3):1–20, 2015.

[3] S. Bennett, J. Van Der Kamp, G. J. Savelsbergh, and K. Davids. Timing a one-handed catch. *Experimental Brain Research*, 129(3):362–368, 1999.

[4] S. J. Bennett, J. van der Kamp, G. J. Savelsbergh, and K. Davids. Discriminating the role of binocular information in the timing of a one-handed catch. *Experimental Brain Research*, 135(3):341–347, 2000.

[5] K. Benzeroual, R. S. Allison, and L. M. Wilcox. Distortions of space in stereoscopic 3d content. In *SMPTE 2nd Annual International Conference on Stereoscopic 3D for Media and Entertainment*, pp. 1–10. SMPTE, 2011.

[6] D. Cheng, Q. Hou, Y. Li, T. Zhang, D. Li, Y. Huang, Y. Liu, Q. Wang, W. Hou, T. Yang, et al. Optical design and pupil swim analysis of a compact, large epd and immersive vr head mounted display. *Optics Express*, 30(5):6584–6602, 2022.

[7] R. D. Cook. Detection of influential observation in linear regression. *Technometrics*, 19(1):15–18, 1977.

[8] W. Epstein. Modification of the disparity-depth relationship as a result of exposure to conflicting cues. *The American Journal of Psychology*, 81(2):189–197, 1968.

[9] Z. Gao, A. Hwang, G. Zhai, and E. Peli. Correcting geometric distortions in stereoscopic 3d imaging. *PloS One*, 13(10):e0205032, 2018.

[10] P. B. Hibbard, L. C. van Dam, and P. Scarfe. The implications of interpupillary distance variability for virtual reality. In *2020 International Conference on 3D Immersion (IC3D)*, pp. 1–7. IEEE, 2020.

[11] J. M. Hillis, S. J. Watt, M. S. Landy, and M. S. Banks. Slant from texture and disparity cues: Optimal cue combination. *Journal of Vision*, 4(12):1–1, 2004.

[12] I. P. Howard and B. J. Rogers. *Perceiving in depth, volume 2: Stereoscopic vision*. Oxford University Press, 2012.

[13] P. Neveu, A.-E. Priot, J. Plantier, and C. Roumes. Short exposure to telestereoscope affects the oculomotor system. *Ophthalmic and Physiological Optics*, 30(6):806–815, 2010.

[14] A.-E. Priot, P. Neveu, B. Priot, C. Prablanc, C. Roumes, and R. Laboissière. Adaptive estimates of egocentric distance from vergence in telestereoscopic viewing: Von hofsten's model revisited. In *2015 International Conference on 3D Imaging (IC3D)*, pp. 1–8. IEEE, 2015.

[15] A.-E. Priot, A. Vacher, C. Vienne, P. Neveu, and C. Roumes. The initial effects of hyperstereopsis on visual perception in helicopter pilots flying with see-through helmet-mounted displays. *Displays*, 51:1–8, 2018.

[16] W. Robinett and J. P. Rolland. A computational model for the stereoscopic optics of a head-mounted display. *Presence: Teleoperators & Virtual Environments*, 1(1):45–62, 1992.

[17] L. B. Rosenberg. The effect of interocular distance upon depth perception when using stereoscopic displays to perform work within virtual and telepresent environments. Technical report, Standford Univ CA Center for Design Research, 1992.

[18] M. Siegel and S. Nagata. Just enough reality: comfortable 3-d viewing via microstereopsis. *IEEE Transactions on Circuits and Systems for Video Technology*, 10(3):387–396, 2000.

[19] R. Spottiswoode and N. Spottiswoode. *The theory of stereoscopic transmission*. Univ of California Press, 1953.

[20] G. W. Stuart, P. Flanagan, and P. Gibbs. Hyperstereopsis in helmet-mounted nvds: slope perception. In *Head-and helmet-mounted displays XII: Design and applications*, vol. 6557, pp. 137–144. SPIE, 2007.

[21] J. Tong, R. S. Allison, and L. M. Wilcox. The impact of radial distortions in vr headsets on perceived surface slant. *Electronic Imaging*, 2020(11):60409–1, 2020.

[22] J. Tong, R. S. Allison, and L. M. Wilcox. Optical distortions in vr bias the perceived slant of moving surfaces. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 73–79. IEEE, 2020.

[23] L. Vaissie, J. P. Rolland, and G. M. Bochenek. Analysis of eyepoint locations and accuracy of rendered depth in binocular head-mounted displays. In *Stereoscopic Displays and Virtual Reality Systems VI*, vol. 3639, pp. 57–64. SPIE, 1999.

[24] H. Von Helmholtz. *Treatise on physiological optics*, vol. 3. Courier Corporation, 2013.

[25] H. Wallach, M. E. Moore, and L. Davidson. Modification of stereoscopic depth-perception. *The American Journal of Psychology*, 76(2):191–204, 1963.

[26] J. P. Wann, S. Rushton, and M. Mon-Williams. Natural problems for stereoscopic depth perception in virtual environments. *Vision Research*, 35(19):2731–2736, 1995.

[27] C. Ware, C. Gobrecht, and M. A. Paton. Dynamic adjustment of stereo display parameters. *IEEE Transactions on Systems, Man, and Cybernetics-part A: Systems and Humans*, 28(1):56–65, 1998.

[28] Z. Wartell, L. F. Hodges, and W. Ribarsky. A geometric comparison of algorithms for fusion control in stereoscopic htds. *IEEE Transactions on Visualization and Computer Graphics*, 8(2):129–143, 2002.

[29] C. Wheatstone. Xviii. contributions to the physiology of vision.—part the first. on some remarkable, and hitherto unobserved, phenomena of binocular vision. *Philosophical Transactions of the Royal Society of London*, (128):371–394, 1838.