ORIGINAL ARTICLE

Contributions of pictorial and binocular cues to the perception of distance in virtual reality

Rebecca L. Hornsey[1] · Paul B. Hibbard[1]

Abstract

We assessed the contribution of binocular disparity and the pictorial cues of linear perspective, texture, and scene clutter to the perception of distance in consumer virtual reality. As additional cues are made available, distance perception is predicted to improve, as measured by a reduction in systematic bias, and an increase in precision. We assessed (1) whether space is nonlinearly distorted; (2) the degree of size constancy across changes in distance; and (3) the weighting of pictorial versus binocular cues in VR. In the first task, participants positioned two spheres so as to divide the egocentric distance to a reference stimulus (presented between 3 and 11 m) into three equal thirds. In the second and third tasks, participants set the size of a sphere, presented at the same distances and at eye-height, to match that of a hand-held football. Each task was performed in four environments varying in the available cues. We measured accuracy by identifying systematic biases in responses and precision as the standard deviation of these responses. While there was no evidence of nonlinear compression of space, participants did tend to underestimate distance linearly, but this bias was reduced with the addition of each cue. The addition of binocular cues, when rich pictorial cues were already available, reduced both the bias and variability of estimates. These results show that linear perspective and binocular cues, in particular, improve the accuracy and precision of distance estimates in virtual reality across a range of distances typical of many indoor environments.

Keywords Virtual reality · Cue combination · Distance perception · Size constancy

1 Introduction

Identifying the distances between oneself and other objects are essential for our everyday actions; from reaching to pick up an object, to perceiving how much leeway there is before stubbing your toe on a table, or cautiously keeping a safe distance from a cliff edge. A broad array of visual information from our environment allows us to interpret where we are, what surrounds us, and what we can do in that space. Cues inevitably vary in both the nature and reliability of information that they provide, and our visual system needs to take this into account in best weighing up the evidence when judging distance. The aim of this three-part study was to determine how visual cues are combined in the perception of distance in complex, naturalistic settings. We used virtual reality (VR) to precisely control the availability of distance

cues and assessed the perception of distance in terms of both its bias and precision using distance sectioning and size constancy tasks.

Distance estimation is influenced by environmental context, the availability of depth cues, and the task for which it is used (Proffitt and Caudek 2003; Wickens 1990). There are many visual cues to depth, and they can be broadly categorized into those that are available via a single monocular image (pictorial cues); those that depend on the differences in the vantage points of our two eyes (binocular cues); those that depend on the motion of objects or the observer (shape from motion and motion parallax) and finally physiological aspects which are guided by distance (such as the accommodation of our lenses, or the binocular convergence of our eyes). In VR, perceived depth may also be affected by *system-related* factors such as the field of view of the display or the accommodation distance required to bring this into focus.

Crompton and Brown (2006) tested egocentric distance estimates in physical space, in industrial and rural settings. Participants estimated the distance to an identifiable

✉ Rebecca L. Hornsey
   rlhornsey@outlook.com

1   University of Essex, Colchester CO4 3SQ, UK

structure and drew increments of this distance on an un-scaled map. It was found that in a rural village, distance was overestimated by a factor of three, compared to an over-estimation factor of 1.5 in the industrialised town. Thus, some environmental factors in the scenery must come into play when interpreting scene distances. The study showed that large biases in perceived distance can occur in the real world and testing such phenomena in VR allows us to keep extraneous cues constant while testing the effect of specific cues in separate, yet consistent, environments. This is the method of the current study.

In real, physical environments, it is difficult to separate and control distance cues while still presenting participants with a rich, naturalistic stimulus. This desired combination of complexity and control is easy to achieve in virtual environments, such as stereoscopic displays and VR (Surdick et al. 1997; Lampton et al. 1995; Barfield and Rosenberg 1995; Witmer and Kline 1998; Hibbard et al. 2017a, b). In VR, it is possible to manipulate the presence of cues and thus assess their importance for the perception of distance. In turn, these findings are important for optimising the display of 3D information in VR. Combinations of cues which are known to improve the accuracy of perceived depth can be prioritised in the rendering of scenes, while those which have a minimal contribution can be omitted. VR headsets place high demands on graphical rendering and it is therefore beneficial for software designers and researchers to establish which visual cues can provide for the best perception of space within these virtual environments, so that power is not unnecessarily wasted on additional components that do not enhance the experience.

Kline and Witmer (1996) suggested that pictorial distance cues such as linear perspective, relative size, brightness, and height in the visual plane have lower importance than physiological and system-related (hardware) cues in near space. An experiment investigating these two types of visual cues was conducted, manipulating specifically the field of view (FoV; system-related) and texture type (pictorial). Participants made distance estimates against a wall in a stereoscopic display, with their head position locked. It was found that FoV had more impact than texture at distances up to 12 ft (3.7 m), and that simply having texture available improved estimates of distance compared to when no texture was present. There was a consistent overestimation of distance in trials with a narrow FoV and consistent underestimation in trials with a wide FoV. The authors concluded that FoV was most heavily relied upon at near distances because the difference in accuracy was greater between the narrow and wide FoVs, than the difference in texture conditions. For instance, the overestimation of distances in the narrow FoV condition was reduced by the presence of texture, but not entirely.

Psychophysical methods have been used to assess the contributions of binocular and pictorial cues on the perception of distance, using stereoscopic displays. For instance, Surdick et al. (1997) used two-alternative forced-choice methods to assess the contributions of multiple cues to the perception of depth. They found that perspective cues were significantly more efficient at conveying depth at a distance of 1 m than cues such as relative brightness or binocular disparity. In the current study, a similar approach is applied in VR, to determine the contributions of binocular and pictorial cues to the accuracy of distance perception under free-viewing of complex, naturalistic displays.

The accuracy of distance perception is characterised by the bias and precision of participants' judgements. Here, bias refers to any systematic errors in judgements, whereas precision refers to the variance of these over repeated judgements. We used two tasks: distance trisection and size constancy, to assess how these measures vary across distance and cue conditions.

Bisection is a commonly used practise for measuring biases in the perception of distance (Rieser et al. 1990; Bodenheimer et al. 2007; Hornsey et al. 2017; Hornsey and Hibbard 2019), in which an observer splits a specified distance into equal halves with a marker. This method allows us to determine whether perceived distance is linearly related to physical distance, or whether there is systematic expansion of compression of far space relative to near space. For example, if far space is perceptually compressed relative to near space, then we would expect the marker to be set closer than the true midpoint, to compensate for this bias. There is a methodological limitation with this method, however, if the random start distance is a uniformly distributed distance between the specified points then even if the marker is not moved during trials then the mean average of this position will closely resemble the true midway point. Additionally, if settings are very imprecise, it is difficult to measure bias accurately. One way to counter these issues is to ask observers to split the distance into more than two sections. This method of trisection was used in Experiment 1. If far distance is compressed relatively to near space, then we expect both markers to be set at a lesser distance. Conversely, if far distance is expanded, both markers would be positioned further away than the correct values. This is presented in Fig. 2.

While distance bisection and trisection tasks can be used to measure nonlinear compression and expansion, accurate settings could be made using this method if distances were under- or over-estimated by the same scaling factor at all distances. To assess biases which might also be a linear scaling of distance, we used a size constancy task (Hornsey and Hibbard 2018; Kopiske et al. 2019; Scarfe and Hibbard 2006; Murgia and Sharkey 2009; Carlson 1960) as an indirect measure of the accuracy of distance perception. In this task, observers indicate the size of an object presented at differing distances. If distance is perceived accurately, the indicated size of a stimulus would be accurate. If the size

settings are too large, and this error is attributed solely to misperception of distance, this would denote the underestimation of distance. Conversely settings that are too small would denote the overestimation of distance. Hornsey et al. (2020) used this method in VR and found that observers' settings were consistent with a progressively greater underestimation of distance with increasing presentation distance consistent with previous psychophysical studies using stereoscopic displays (Johnston 1991; Brenner and van Damme 1999; Scarfe and Hibbard 2006).

Here, four environments were created with increasing amounts of pictorial depth cues as shown in Figs. 1 and 3. In the simplest condition, the *Sparse* scene merely presents a floor plane. In this condition, the principle cues to distance are binocular convergence and the size of the image of each object. Linear perspective was added by presenting walls and a ceiling to create the *Perspective* scene. The *Textured* scene added bricks and a floor pattern in order to include texture as a cue to distance. Finally, in our *Cluttered* scene, object clutter was added to the scene to provide a broad range of pictorial and binocular cues. The step from Textured to Cluttered added potential anchors into the scene which were able to provide supplemental depth cues through additional disparity information, or the integration of information across distance (Wu et al. 2004). Evidence suggests that texture might be the most important cue for distance perception: Sinai et al. (1999) found that estimates were better when brick floors were added to a stimulus, and Thomas et al. (2002) found depth matching estimates were more accurate with vertical line textures visible.

Kline and Witmer (1996) determined that the most accurate perceptions of virtual distances arose from having a wide FoV and a pattern with high resolution. In addition, Loomis and Knapp (2003) hypothesised that any distance compression found in VR might be due to the simplicity of virtual environments, hence the final environment used in the current study incorporated an array of naturalistic objects to the scene.

In the first and second experiment, distance perception was assessed in each of the four scenes using distance trisection and size constancy tasks, respectively. In Experiment 3, a repetition of the Sparse and Cluttered scenes from Experiment 2 (Fig. 3) were used with both monocular and binocular viewing, to assess the importance of binocular cues to the perception of distance under free viewing. The viewing distances tested were across a range representative of many indoor environments.

In addition to measuring systematic bias in the perception of distance, we also measured the precision of observers' settings. For each observer, this was quantified as the variance of settings across trials (σ^2). According to the weighted averaging model of cue combination, perceived distance is an average of the estimates of distance derived from multiple individual cues. In our case, we can predict how precision is improved by the weighting of pictorial and binocular cues (Murray and Morgenstern 2010; Landy et al. 2011):

$$\text{Cues Combined} = W_p D_p + W_b D_b \qquad (1)$$

where W_p and W_b are the weights associated with pictorial and binocular cues, and the distance estimates from these
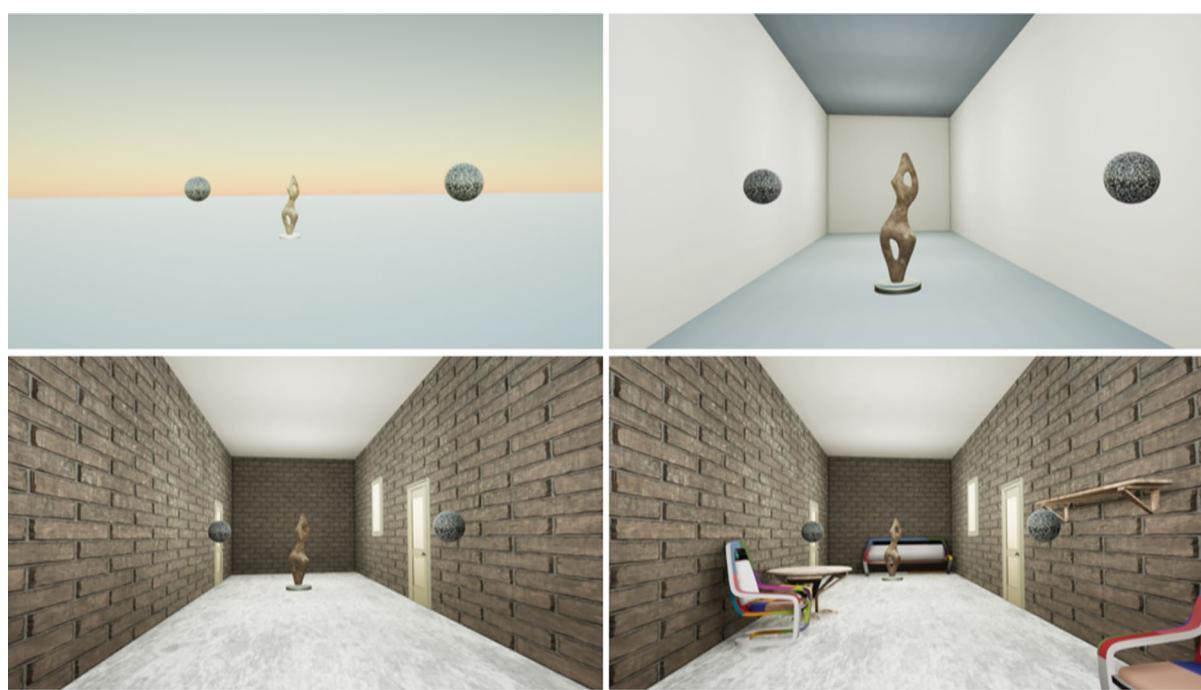
Fig. 1 The four trisection task environments: Sparse (top left), Perspective (top right), Textured (bottom left), Cluttered (bottom right)

cues are given by D_p and D_b, respectively. These estimates are assumed to be unbiased. Averaging the cues with these relative weights, which are proportional to their inverse variance, will optimally improve the precision of the resulting estimate:

$$\sigma^2 c = \frac{\sigma^2 b * \sigma^2 p}{\sigma^2 b + \sigma^2 p} \tag{2}$$

This experiment directly measures the degree to which the precision of distance estimation is improved by the availability of binocular cues. Overall, the aim of this series of experiments was to establish how binocular and pictorial cues contribute to the accuracy and precision of distance judgements in virtual reality and to assess the contributions of these cues against the prediction of a weighted-averaging model of cue combination Landy et al. (2011).

2 Experiment 1: methods

This experiment used a distance trisection task in environments with increasing pictorial depth cues. This allowed us to determine how these cues contribute to the accuracy of distant estimates.

2.1 Participants

A total of 40 participants were recruited from the University of Essex's online participant advertisement system (SONA). The ages ranged from 18 to 26 with a mean of 20. Participants were rewarded with course credits, completion time averaged approximately 60 minutes.

2.2 Stimuli

The reference stimulus (a statue obtained from Unreal Engine's Starter Content) was used to indicate the distance to be trisected. Two spheres were set 50 cm either side of the reference stimulus and were referred to as the *far ball* and *near ball* (both *target stimuli*). These were presented at eye height.

The other cues in the room were loaded from Unreal Engine starter pack, where each environment had specific cues visible:

– *Sparse* where only a floor surface was visible. Distance information is provided primarily by the size and binocular cues of the reference and target stimuli themselves.
– *Perspective* with the addition of plain surfaced walls and ceiling. The additional distance information here is provided by the linear perspective from the intersections between the walls, ceiling and floor.

– *Textured* a brick pattern on the wall and a textured floor covering provide additional texture cues.
– *Cluttered* furniture and shelving created a more cluttered environment, providing a rich, multi-cued environment to provide a natural gamut of depth cues.

The colour scheme and light sourcing were rendered to provide a naturalistic setting, with the same static light illuminating all environments. Each environment is shown in Fig. 1.

2.3 Apparatus

A PC with an Nvidia GeForce GTX 1060 graphics card was used to create and present the stimuli, through Unreal Engine 4.18. The Oculus Rift was used to present the environments, with an adult population average inter-ocular distance of 63 mm (Dodgson 2004) set for all participants. Participants used the Oculus Motion Controller to take part in the experiment. These apparatus were used for all experiments.

2.4 Task and procedure

In each environment, the reference and target balls were constantly visible. Target stimuli were set to random X axis (forwards/backwards egocentric distance) locations between the participant and the reference stimulus between each trial. Participants had control over the X axis position of the target balls through button presses on the matching side left/right controller. The aim was for participants to position the near ball at one-third of the distance between theirself and the reference stimulus and the far ball at two-thirds of the distance.

After instructions on how to complete the task were given to each participant, they moved a chair into a specified start position which was indicated by a pink circle on the floor in the virtual environment. The reference stimulus was positioned at distances of 3, 5, 7, 9 or 11 m, varied between trials. Once the participant had decided upon the distances of both target stimuli, they indicated this to the experimenter so the absolute distances of the targets could be saved and to move onto the next trial.

The order in which the participants completed the environments was counterbalanced, as was the distance of the reference stimulus within them. Each block of trials consisted of 15 repetitions for each of the 5 reference distances (3, 5, 7, 9 and 11m). These 75 trials were presented in a randomised order.

3 Experiment 1: results

A linear mixed effects model was used to predict the distance settings, plotted in Fig. 5. For each environment there were two sets of analyses conducted, with either

intercept only or intercept and reference distance (D) as random factors across participant (P). The model that we used predicted the distance setting using the reference distance as a fixed factor and a random intercept and slope:

$$\text{Distance Setting} \sim D + (1 + rD|P) \tag{3}$$

The results of fitting this model are shown in Table 1. All near third slopes, in each environment, were significantly different from 0.333 at the $p < 0.05$ level and for the far third slopes all bar the Sparse environment produced values significantly different from 0.666. These are presented in Fig. 7. The intercepts for the near thirds were all significantly larger than zero; all far third intercepts bar the Sparse environment were also significantly larger than zero.

The accuracy of distance settings, which are presented as the mean signed error in Fig. 4 shows the near third settings were consistently set closer than the true first third distance in all environments, while the far third settings were overshot in the Sparse and Perspective environments, but also undershot in Textured and Cluttered conditions. A one-way ANOVA was used with environment as a predictor for unsigned error [near $= F(3, 8994) = 6.991$, $p < 0.0001$; far $= F(3, 8994) = 19.615$, $p < 0.0001$] which led to repeated t-tests being carried out: for near settings a significant difference between Sparse and Perspective environments ($p < 0.0001$) was found and for far settings there was a significant difference between Sparse and Perspective ($p < 0.0001$) and Textured and Cluttered environments ($p < 0.0001$).

The average of the standard deviation for each participant's repeated settings is presented in Fig. 6. The most noticeable decrease in variation occurred between the Sparse and Perspective environments. In order to normalise the data, the fourth root of the standard deviation was calculated for use in all precision analyses (Hawkins and Wixley 1986; Todd et al. 2010). The analysis was a mixed effects linear model, using distance as a predictor and participant as a random structure on both the near and far settings separately, across all four environments. For the near-setting data set, both intercepts and slopes were significantly different from zero at the 0.01% significance level which led to post-hoc analyses to compare the improvement of precision in the Cluttered environment to all other environments and to further explore the interaction of distance and environment. The improvements from the Sparse ($p = 0.014$) and Perspective ($p < 0.0001$) environments were notable, as was the difference between the Sparse ($p = 0.010$) and Cluttered ($p < 0.0001$). The same analyses were conducted on the far-setting data revealing the same significant main effect of environment on the precision, and the follow-up regression main effect of condition improvement was significant in the Sparse

($p < 0.0001$) and Perspective ($p = 0.019$) also interactions of condition and distance within Sparse ($p = 0.040$) and Cluttered ($p < 0.0001$) environments.

4 Experiment 2: methods

In this experiment, a size constancy task was used with the same environments assessed in Experiment 1.

4.1 Participants

A new group of 20 participants were recruited from the University's online participant recruitment system, who had a mean age of 21 and ranged between 18 and 27.

4.2 Stimuli and apparatus

The same four environments from Experiment 1 were used, but instead of the target and reference stimuli there was only one target sphere. The target axes were locked in that when the size was manipulated either by the participant or between trials, it would increase/decrease as a sphere. A size 5 (22 cm diameter) football was used as a physical reference for participants.

4.3 Task and procedure

Participants had the headset's controller in one hand and the football in the other/on their lap. Before starting, they moved into the specified start position and were instructed not to move the chair from this position.

In each trial, a target stimulus was visible at each of the previously tested egocentric distances of 3, 5, 7, 9, and 11 m. Controller buttons were used to increase/decrease the size of the target object until the size appeared to match the size of the football participants held. Participants verbally indicated they had completed the trial to the experimenter, at which point the metrics were saved and moved onto the next trial or new environment.

The order of environments was randomised, as were the distances of the target and the size of the target in each new trial.

5 Experiment 2: results

Figure 4 shows the decrease in size errors as more environmental cues were made available. Variance also reduced as pictorial cues were increased (Fig. 6). Constancy errors were lowest in the full cue condition, although when converted into the effective distance, the slope value from this and all other environments was nevertheless significantly lower

than one (accurate performance), Fig. 8, indicating a lack of perfect constancy in size perception across distance.

The results for accuracy, as shown in Fig. 4, were analysed using a one-way ANOVA with condition as a predictor for the unsigned error. The significant output [$F(3, 2994) = 79.172$, $p < 0.0001$] led to repeated t tests to be carried out; significant differences were found between all environments ($p < 0.0001$), with the Sparse environment presenting the lowest accuracy. Accuracy increased consistently with the addition of the tested visual cues.

Using this data to indirectly measure the distance perceived to the stimuli, the data points were converted to the effective distance perceived in each trial. These new data points are also shown in Fig. 5 with accurate performance displayed as the black line. Another linear mixed effects model analysed these new *perceived distance* data in each environment, including target distance (D) as a fixed effect, and target distance as a random slope; this reduced the AIC compared to a model in which this random slope was not included:

$$\text{Perceived Distance} \sim D + (1 + D|P) \qquad (4)$$

The output values in Table 2 from Eq. 4 show that in each environment, all slopes were significantly different from one (accurate performance) at the $p < 0.05$ level. Figure 7 depicts these slopes for each environment with 95% confidence intervals. All distances were perceived as being significantly less than what was rendered across each condition, with the most compression found in the Sparse environment and least in the Cluttered environment.

The standard deviation and 95% confidence limits are shown in Fig. 6, with the greatest variability found in the Sparse environment. The fourth root of the standard deviations was calculated and centred for analysis. Mixed effects models to predict these values (in each environment separately) using distance as a predictor and participant as a random structure revealed significant differences at the 1% level, showing that the variance in settings was consistently altered by presentation distance across environments. This led to post-hoc analysis of the distance * environment interaction and to compare the Cluttered environment's improvement against the other environments. No significant interactions were found, although the improvements of precision in the Cluttered environment were significantly different from the Sparse and Perspective environments (both at $p < 0.0001$).

6 Experiment 3: methods

This was a replication of Experiment 2, using only the Sparse and Cluttered environments, however, each participant viewed the two scenes both monocularly and binocularly. This allowed us to directly test the contribution of binocular cues to improving the accuracy and precision of distance perception in VR across a range of distances.

6.1 Participants

Another 20 participants were recruited from the University's advertisement system, who ranged between 18 and 26 and had a mean age of 20.

6.2 Stimuli and apparatus

The Sparse and the Cluttered environments from Experiment 2 were used. The conditions were:

– *Monocular sparse* viewing through one eye, only a floor surface was visible
– *Monocular cluttered* naturalistic corridor components through one eye
– *Binocular sparse* the floor surface through both eyes
– *Binocular cluttered* corridor components through both eyes

The same size 5 football was used as a physical reference for the task. When in the monocular conditions, tissue was used to cover the left lens of the headset.

6.3 Task and procedure

The same start position was used for participants; holding both the controller and the football in each hand. The two face buttons on the controller increased/decreased the size of the virtual sphere and when participants had matched the size of this to the size of the physical football the experimenter manually saved the measurements and moved the experiment onto the next trial.

The virtual sphere was presented at 3, 5, 7, 9, 11 m in front of the participants. The start size of the virtual sphere was randomised between each trial, the order of presentation distance was also randomised, and the condition order was counterbalanced.

7 Experiment 3: results

As can be seen in Fig. 4, adding environmental cues (clutter) to the Monocular sparse condition reduced the mean size errors by 18.66 cm, and adding binocular disparity to the sparse environment decreased size errors by 8.24 cm. The addition of these two cues also influenced the precision of participants' size estimates, shown in Fig. 6, whereby the amount of variability in the Binocular cluttered environment

is significantly (15.09 cm) less than that in Monocular sparse.

The same regression analyses ran in Experiment 1 and 2 were conducted on the perceived distances for each environment, again with the best AIC resulting from including reference distance as a random slope (Eq. 4). In Table 3, it can be seen that only the intercept value for the Binocular cluttered environment is *not* significantly different from zero, showing most accurate performance in this condition. Slopes in all conditions are significantly lower than one, as plotted in Fig. 7, showing again incomplete constancy, but best performance once again in the Binocular cluttered environment.

A one-way ANOVA using environment to predict the amount of unsigned error produced a significant effect: $F(3, 2994) = 78.409$, $p < 0.0001$. Follow-up t-tests revealed significant differences between all environments at $p < 0.0001$, showing that the Binocular cluttered environment achieved the highest level of accuracy.

After centring the data, a significant (at the 0.01% significance level) categorical mixed effects model was conducted to compare the fourth root of the standard deviations of each environment to that of the Cluttered results. This was to check if the change in precision occurred only at some distances for any environment and revealed all interactions were significant with Monocular sparse at $p = 0.050$, Monocular cluttered at $p = 0.013$, and Binocular sparse at $p < 0.0001$. The intercept improvements from all environments to the Binocular cluttered were also significant at the 0.01% level.

A final t-test was conducted to compare the predicted precision of the full cue condition to the actual results. The fourth root of the standard deviations for each environment was used to obtain the predicted standard deviation, which was calculated as:

$$\sigma_P = \frac{\sigma_{MC} \cdot \sigma_{BS}}{\sigma_{MC} + \sigma_{BS}} \tag{5}$$

where σ_P is predicted standard deviation in the full cue condition, σ_{MC} is the standard deviation in the Monocular cluttered condition, and σ_{BS} is the standard deviation in the Binocular sparse condition. The reduction in standard deviation was significantly lower than predicted, $t(99) = -5.658, p < 0.0001$: precision was better than expected in the full cue environment tested here.

8 Discussion

The effects of the availability of pictorial and binocular cues on the accuracy of distance perception in VR were assessed using trisection and size constancy tasks. The trisection experiment was a direct measure of distance perception,

whereas the constancy experiments were indirect. We assessed the bias and precision of distance perception using these tasks.

8.1 Bias and precision

Each of the experiments tested the effect specific visual cues had on performance, which was measured in terms of accuracy and precision. In relation to this, however, there is a distinction between precision and bias that is not always taken into consideration when using cue-combination models. Through the analysis procedure used here, a bias of (linear) underestimation has been revealed: in Fig. 7 the regression values all less than one indicate this. As well as this, Fig. 9 shows the standard deviation varying across distances and environments which instead demonstrate the metric of precision.

In studies of cue combination, it is often assumed that there will be no bias arising from either of the two factors being tested (here: distance, environment) along with the assumption that the factors will create random variation (Scarfe et al. 2011). In this instance, however, it has previously been established that distance perception is biased in the direction of underestimation. However, it was still predicted that precision would improve if cues can be combined through appropriate weighted averaging.

In Experiment 3, the precision in the full cue condition can be predicted from the data in reduced cue conditions using Eq. 2: the contribution of binocular disparity *and* pictorial cues was able to be predicted for the Binocular cluttered environment. By using the values from Fig. 6 (Monocular cluttered = 15.682; Binocular sparse = 15.343), Eq. 5 gives the predicted standard deviation of 9.711, CLs $[-4.192, 23.615]$ if the two sets of cues were combined optimally. The results give an actual standard deviation of 6.934, CLs $[-8.486, 22.354]$, which are consistent with the preicted improvements from cue combination, as this result is within the credible range of the wide estimated standard deviation.

The findings of the current study are consistent with prior research into the contribution of stereoscopic viewing: the binocular cues added when viewing a scene stereoscopically can improve sensitivity to depth (Hornsey et al. 2015). It has been concluded that binocular disparity provides complimentary depth information to the monocular pictorial cues. Just like the findings in the current study, when combining the two types of cues the performance is significantly enhanced to be greater than any one subset of cues alone.

A study by Kim et al. (1987) compared performance in a task using a monocular and a stereoscopic display: it was found that incorporating a texture grid with ground intercepts produced performance that was equal to that of the stereoscopic display alone. This is similar to the findings of Experiment 3 (Fig. 8), that the distances perceived from

monocular viewing with full pictorial cues were uniform with distances perceived binocularly but with no pictorial cues. Using this finding for future work with similar equipment, or wanting to get the same desired outputs, developers should only use cues which aid 3D vision in stereo displays as it will reduce complexity of computer system. If costly cues like binocular disparity are just as effective as simple pictorial cues then we should go with the least costly option.

8.2 Comparing to the hypotheses

The first hypothesis, that there would be an underestimation of near space compared to far, in Experiment 1 could be shown by either the near third position consistently set too close but with an accurate far third or setting both thirds too close when the reference stimulus is at the closer distances but accurate setting at the further distances tested. This systematic underestimation of the near third (of around 8 cm) is found in the results of Fig. 4 pointing to the underestimation of near space relative to far. In Experiments 2 and 3 this would be portrayed by size settings becoming larger as the presentation distance increases. Figure 5 indeed shows that overall, there was an underestimation of near space as an increasing amount of the size settings are above the black line of accurate performance as presentation distance becomes greater. Following this up with statistical analyses on the impact of presentation distance on perceived distance (derived from size set): it can be seen in Fig. 7 that in the constancy tasks, the effect of presentation distance was reduced when visual cues were added, as the slope gradient for each environment's regression approaches one (whereby perceived distance is equal to presentation distance).

The second hypothesis, that performance would improve with the addition of cues, was also supported by the results from the constancy experiments. In Table 2, the slope values increase with each cue that is added (accurate performance slope = 1). The values are still less than one, indicating incomplete constancy. Accuracy, measured by signed error, is shown in Fig. 4 to improve with each additional cue in Experiment 2 and the addition of context and disparity separately in Experiment 3. Experiment 1 only showed improvement of accuracy from the addition of visual cues from Sparse to Perspective.

The other aspect of performance was measured by the standard deviation of settings across participants in each condition and distance. In Fig. 6, each graph shows how the variability in participants' performance differs with the addition of cues. The overall standard deviation values decrease with the addition of cues which is a result of a higher sensitivity to distance in each trial. Taking the results for both accuracy and variability into consideration, we have concluded that performance is indeed enhanced with the addition of these specific visual cues.

Experiment 3 replicated the Sparse and Cluttered environments with binocular and monocular viewing, to determine the contribution of binocular disparity compared to pictorial cues. It was hypothesised that higher performance would be obtained by binocular than monocular viewing in the Sparse condition and that the Cluttered environments would increase performance than the Sparse environments. The t-test analyses found evidence for binocular viewing significantly improving the perceived distances, seen in Fig. 8. The second part of this hypothesis was that there would be less of a difference found between the Monocular cluttered and Binocular sparse conditions, which was indeed found to be a non-significant change in both the precision and accuracy tests.

8.3 Interpretation of findings

Performance was worst in the Sparse environments in each experiment, shown by the combination of inaccurate slope and intercepts in Tables 1, 2, and 3. Murgia and Sharkey (2009) also found that distances were underestimated in both poor and rich cue conditions, but there was greater underestimation in the poor cue environment. Here, the introduction of more pictorial cues aids participants' perception both in the form of precision and accuracy of settings. In Experiment 1, the addition of each positively impacted the precision, and in Experiment 3 the addition of context and disparity had an extremely positive impact on both accuracy and precision. However, Surdick et al. (1997) suggested that linear perspective and texture gradient are among the most useful cues for distance perception. In Experiment 2, the set size of the target is most level across distances in the Texture environment (slope value nearest one in Fig. 7), however, the confidence intervals for the slope value of the Texture environment do not include one, indicating significantly underconstancy. The low intercept of 16 cm obtained from this environment in Table 2 suggests that texture is the best cue for distance estimation at the very close distances, although the slope value is still just under half of what it should be, meaning the stimuli here are still mis-perceived. From these differing results, it could be concluded that the addition of these specific cues is task-dependent for assisting in minimising response accuracy, but there is great evidence that this has a positive effect on minimising response variability.

In addition to this, the findings of the current study support those of Livingston et al. 2009, who investigated whether the perspective cues found indoors, such as the linear alignment of the floor-walls-ceiling, improve outdoor distance perception when using augmented reality. An underestimation of distance was found in the indoor environment, much similar to the findings here. In general, perspective cues require a ground intercept, which is why the baseline condition presented this initially as opposed to the

stimuli floating in space. From this, each additional cue was easily added to the previous environment.

The headset used here has a fixed accommodation distance, unlike in physical space where accommodation is a distance cue that changes according to the fixation distance. A set-up by Bingham et al. (2001) found overestimates in a limited cue experiment, and when the focal plane was reduced by 2 diopters, the overestimation was halved. One might suggest this factor could be influential over the perception of the surroundings in these environments, however research has suggested that accommodation does not have much of an impact when in full-cue conditions (Mon-Williams and Tresilian 1999, 2000). In addition, the cue of accommodation becomes less reliable with age (Ellis and Menges 1998), so given the spread of ages used in the three experiments here, the constant accommodative distance is less likely to have an impact as other factors.

It could be argued that the indoor environments used here might limit the applicability of the findings: Lappin et al. (2006) collected depth judgements from open outdoor environments to find that they were nearly veridical, but indoor environments produced an expansion of perceived distances and more variability in responses. However, Bodenheimer et al. (2007) expanded on this with different distances (15 and 30 m), with compression being perceived only at the far distance, and found no significant difference for indoor or outdoors. Future research replicating the current methodology and equipment in an outdoor setting would be useful in identifying if there is a difference caused by having walls and a ceiling indoor, which emphasise the cue of linear perspective.

Perception of distance from all pictorial cues is equal to the perception of distance from binocular disparity alone (Experiment 3, Fig. 8). This is consistent with findings of certain studies (Eggleston et al. 1996; Roumes et al. 2001; Creem-Regehr et al. 2005); however, there is evidence for the reducing effect of binocular disparity with increase of presentation distance (Hornsey et al. 2020), so is it that there is also a decrease in effectiveness of pictorial cues? Further investigation into this would be useful to identify the effects

Fig. 2 Relationships between actual distance and perceived thirds in the trisection task, presented for two possibilities of nonlinear misperception of distance. Accurate linear perception in black; perceived thirds to be set not far enough in the expansion instance; perceived thirds shown to be set further than the true thirds in the expansion instance
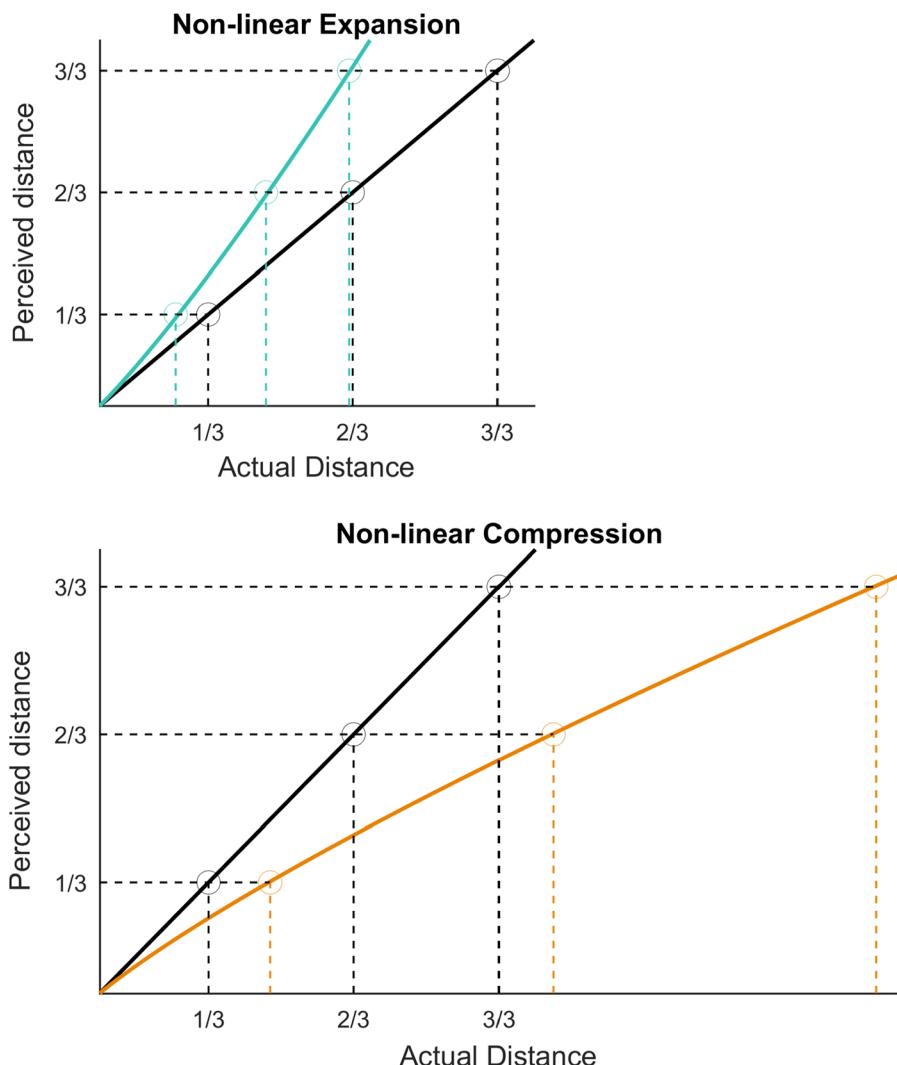
of both at even further distances, and if the relationship between them remains equal.

8.4 Possible explanations of findings

The accuracy results of Experiment 1 show relatively small errors (average of 10 cm in either direction), but this does not necessarily mean that participants were perceiving the rendered distances correctly. Apparent size of objects decreases as a function of distance between it and observer: at far distances changes in apparent distance are less apparent than at closer distances (Sedgwick 1986). The near and far target balls in our experiment were the same size meaning that participants could have compared the size so that the near would be double the size of the far ball. This could be an explanation for the results, however, even by doing this, participants would still need to set the distance of one of the balls correctly. Because of this, a more plausible explanation is that the relative distances between the stimuli were being perceived accurately. In other words, an observer may section the 9 m between theirself and the reference stimuli equally by positioning the near target stimuli at 3 m and the far at 6 m, but they might be perceiving the total distance as 30 m and positioning the target stimuli at 10 and 20 m. This better fits the findings, as the amount of error, in centimetres, in the constancy experiments show the deficit in perception of absolute distance.

In both the Sparse and Perspective environments of Experiment 1, the far thirds were set further away than where they should have, at about the same magnitude of the under-setting of the near thirds. This is indicative of near space expansion and far space compression. Interestingly, this shifts to an overall compression of space in the Textured and Cluttered environments. A possible explanation for the perception of half-distance could be that an overload of visual information of familiar objects with accurate physical properties causes confusion for the sizing of the floating target stimulus. It is most apparent in the Cluttered environment that the target object does not *belong* in the environment like the other objects, so this might impact on the ability to judge its distance relative to the context. It could be argued that this is the most realistic out of all the environments so that is the reason why performance is worse there when the target does not fit in: Singh et al. (2010) found that the presence of a highly salient physical surface does have an effect on distance perception, but not in a systematic way. This switch to a generally consistent, yet inaccurate perception of the space in the Cluttered environment, from a distorted perception in the Sparse ones could be a clue as to the specific cues which are favoured by the visual system.

An alternate explanation for the misperception of distances seen in Fig. 8 is that at specific distances there may be certain cues which are ignored. A study by Glennerster et al. (2006) found that observers ignored available stereo
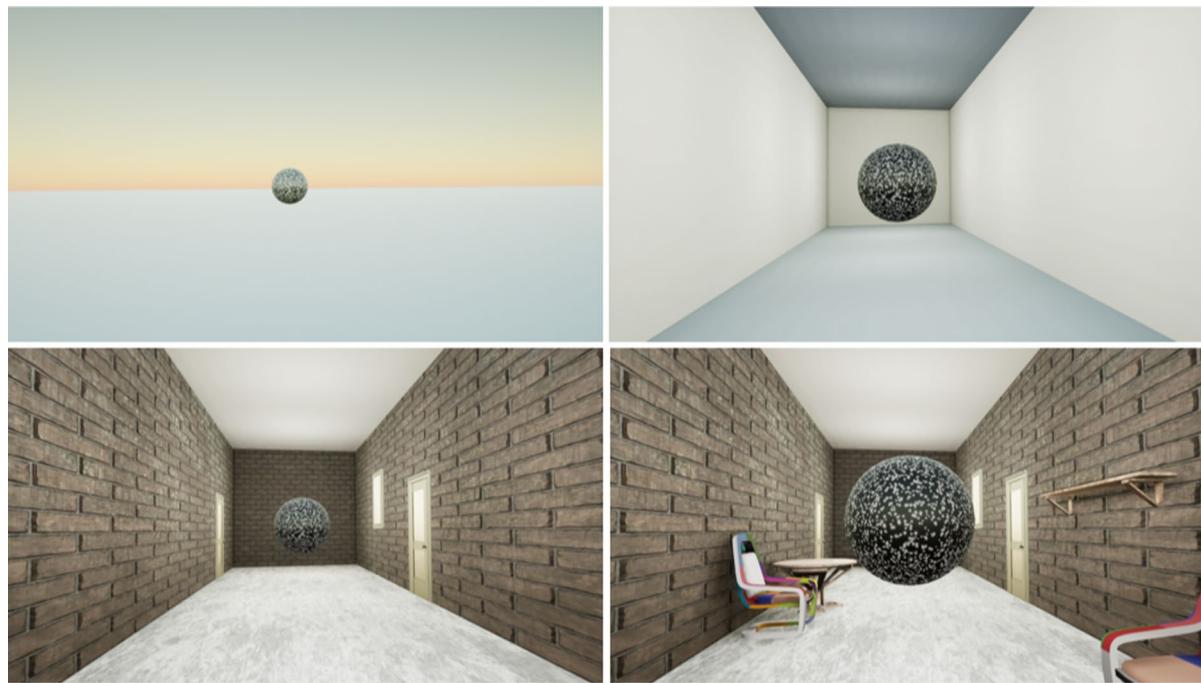
Fig. 3 The four constancy task environments: Sparse (top left), Perspective (top right), Textured (bottom left), Cluttered (bottom right)

Fig. 4 Mean signed error for all environments in each experiment, with 95% confidence limits. Accurate performance would have zero mean error, negative values indicate undersetting of participant response
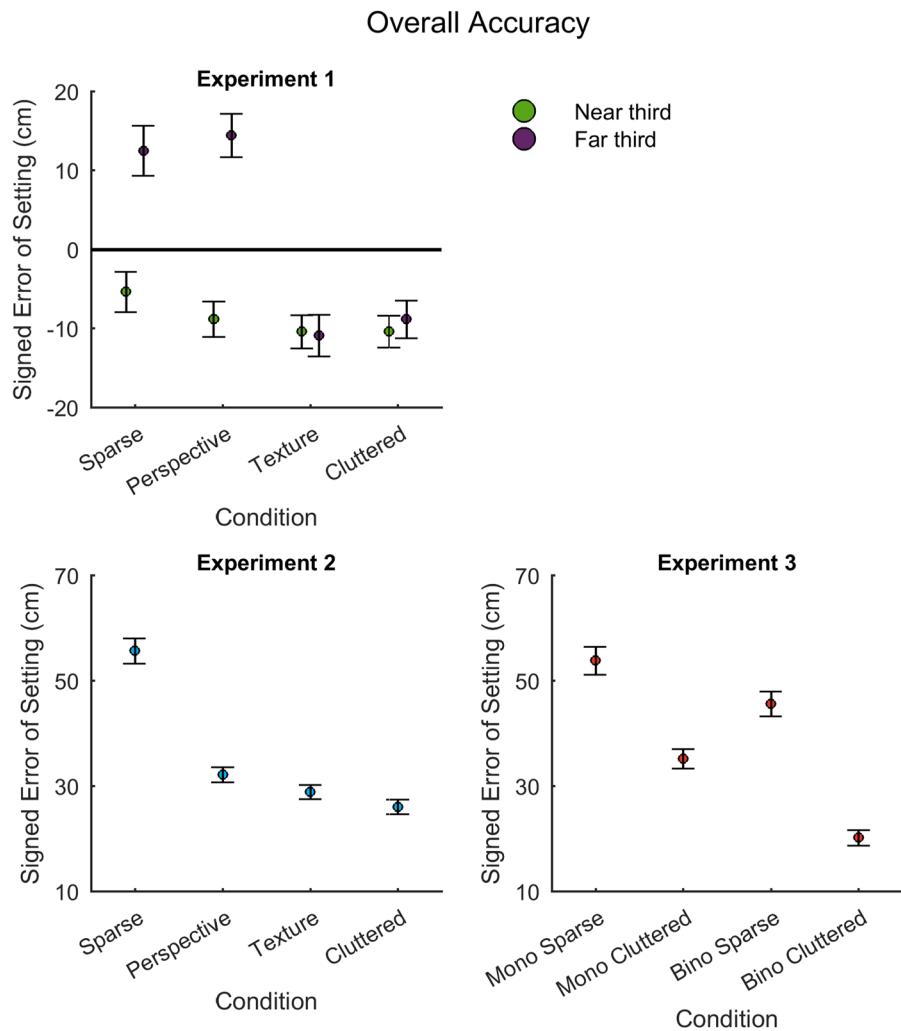

and motion cues when making distance estimates in an immersive HMD and relied on differing cues more heavily at differing distances. It was found that as viewing distance increased from 1.5 to 3 to 6 m, observers used information from the surrounding clutter objects (such as relative size) as anchors for their estimates, rather than the information from binocular disparity. These findings were taken from a setup with a much more limited distance range than what has been presented here, however, the explanation may be relevant: at further distances binocular disparity would be ignored and pictorial cues more heavily relied upon and this would be represented by more variable responses at further distances in both binocular environments, along with more stability in the estimates of the monocular environments. In Fig. 9, the trend in precision over all distances suggests the most precise estimates are at closer distances, compared to far. The effects or pictorial and binocular cues are approximately summative at 3 m during Experiment 3, in that the variation shown in Binocular cluttered is the sum of both these conditions separately:

they are equally as effective at enhancing perception. At around 9 m, they diverge and then at 11 m the improvement is being driven almost entirely by monocular cues. This could be argued to be coinciding with the nonlinear compression prediction in Fig. 2, however, the trends within the raw data plot of Fig. 5 in fact present a linear relationship between perceived and presented distance. Perhaps the reason, the nonlinear relationship found by Glennerster et al. (2006) was not found here was due to the difference in paradigm. Their method intended to instigate a cue conflict between pictorial and binocular information, whereas that was not done here.

A potential cause of the degree of variability in responses throughout each of the three experiments might be from the distance of the two cameras within the headset being kept the same for every participant, despite participants not all having the same distance between their eyes. Drascic and Milgram (1996) concluded that even small alterations in interpupillary distance can cause large distortions. Future research may want to record participant's

Fig. 5 Raw data for each participant in a different colour across all trials. Each row for each experiment with accurate setting in black. One data point was excluded from the final graph for visual consistency with other graphs, but it was included in the analysis
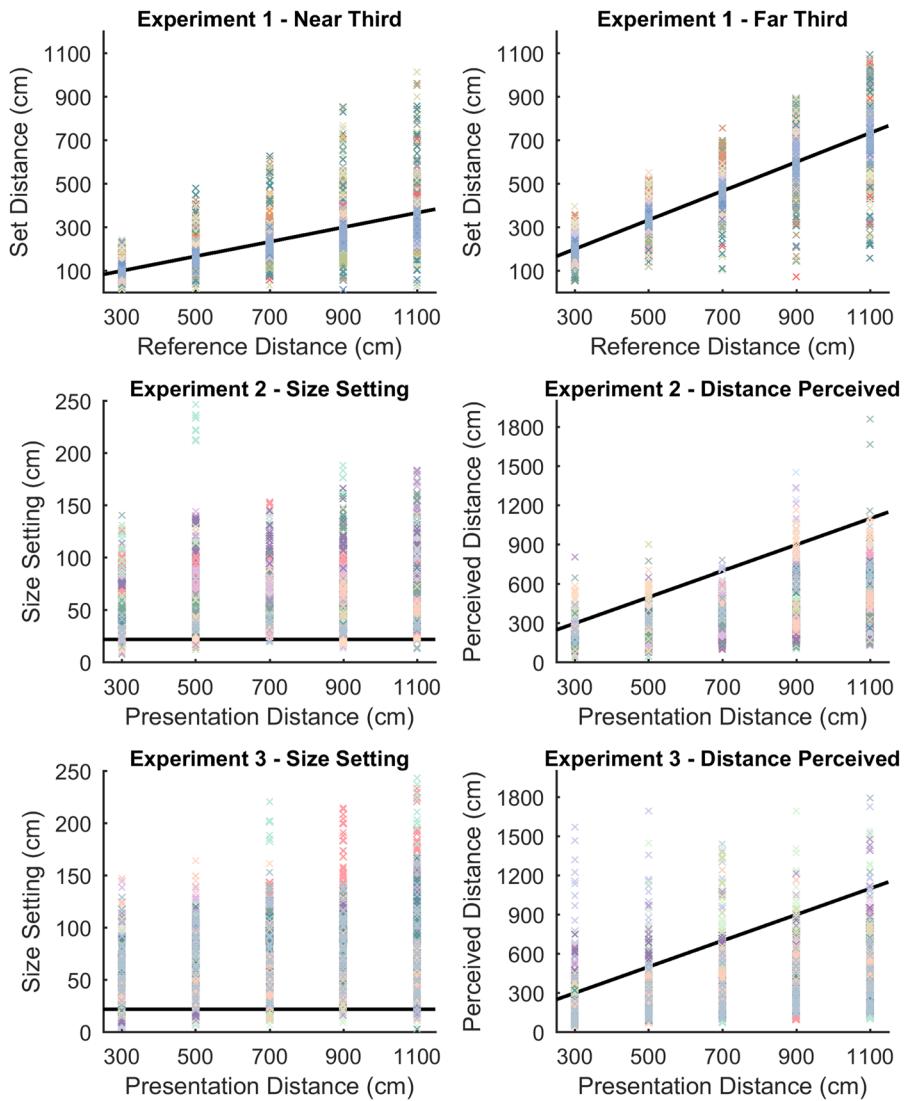

interpupillary distance, not necessarily to change the headset, but to identify any relationship between this and performance. An additional explanation for the amount of variability across trials could be due to the virtual balls not being attached to the floor: when an object floats above the bottom plane, the relative height and distance from the observer becomes more difficult to interpret (Drascic 1991; Kim et al. 1987). For this reason, the tasks used here may have produced results which may be more variable than if the stimuli were attached to the floor plane, much like what would be observed in physical settings.

Some participants may have understood the task straight after viewing the first environment, however, others may have needed extra time to fully get to grips with the task at hand: Drascic (1991) tested stereoscopic and monoscopic displays to find that stereoscopic displays enhanced initial learning of the task. This performance advantage existed even after large amounts of practice for some participants.

If this was the case in Experiment 3, those who were better able to understand the task from doing a binocular condition first, compared to half of the participants who did the monocular condition first, would have been filtered out, as the order of conditions were counterbalanced.

Renner et al. (2013) concluded that in order to enable the best possible perception of distances, quality graphics, binocular disparity, a strong sense of presence, and a texture-rich environment are essential. In the current study, it can be argued that binocular disparity and a texture-rich environment were the only aspects which were directly addressed. The presence was not measured, and the graphics of the headsets were not nearly as high as the capabilities that ultra-HD screens and advanced CGI allow for. Perhaps, increasing these two features may lead to a higher degree of accuracy: an experiment using the same procedure but with more advanced technology with better

Overall Precision

Fig. 6 The amount of variation in responses calculated from the average standard deviation for each participant at each distance, plotted for each environment. Ideal performance would have zero variance, 95% confidence limits as error bars. In the trisection task first showing the near third (green) then far third (purple) setting. Experiment 3 shows variation prediction of cue combination model in mint (color figure online)
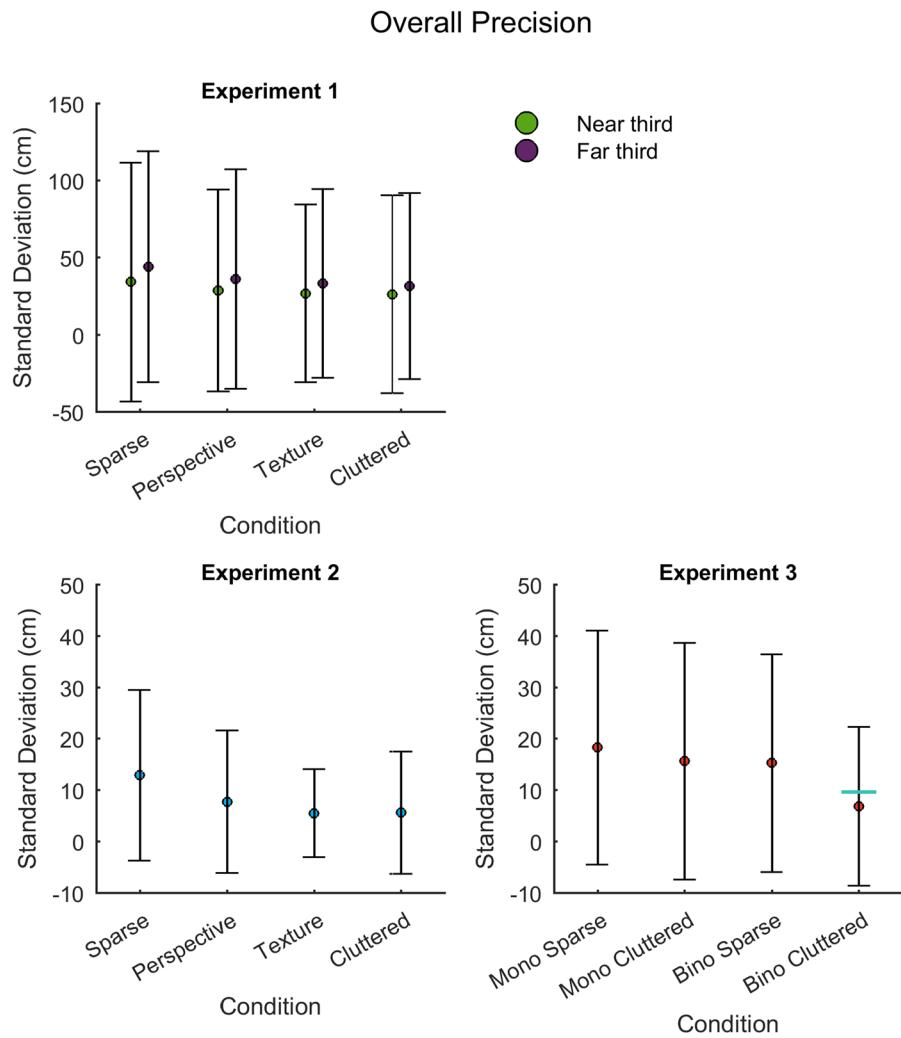
rendering capabilities and also a measure of the presence may produce more accurate results.

8.5 Conclusion and impact

Overall, the results show systematic underestimation of distance, as well as the accuracy and precision of distance perception improved with the addition of both pictorial and binocular cues. For pictorial cues, precision and accuracy both improved most with the addition of linear perspective in the scene. Binocular disparity and pictorial cues were both found to enhance distance perception. The benefits of binocular disparity, in particular, are important to note. Bias and variability were both reduced by the presence of binocular cues, across a distance range typical of many indoor environments. This improvement illustrates the importance of binocular cues, given the costs of the additional resources required to provide binocular information and the additional

Fig. 7 Regression slope values for distances perceived in all environments in each experiment, with 95% confidence limits. Accurate setting indicated by the black lines (trisection experiment-near third 0.333 and far third 0.666 are dashed; constancy experiments 1)
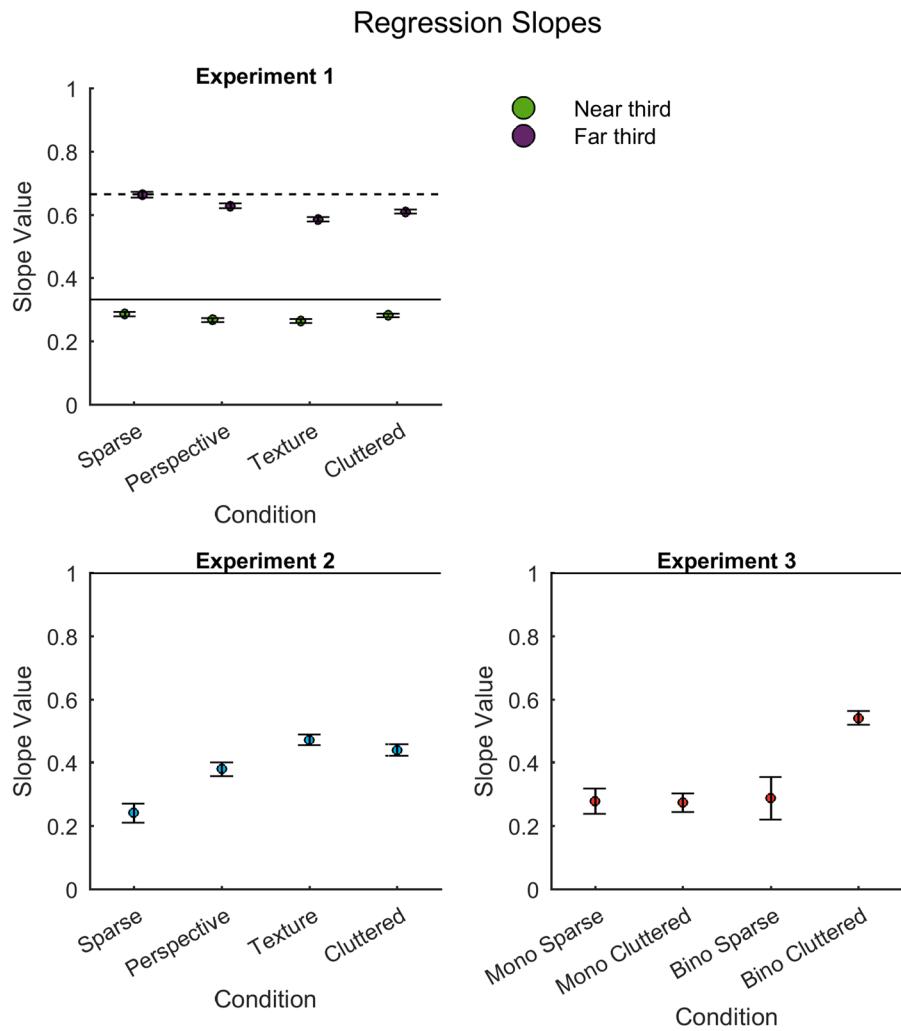

Table 1 Regression analyses of each environment in Experiment 1 from Eq. 3

Environment		Intercept estimate	Intercept CLs	Slope estimate	Slope CLs
Sparse	Near	26.390	[15.195 37.585]	0.288	[0.267 0.309]
	Far	14.385	[− 7.541 36.311]	0.664	[0.626 0.702]
Perspective	Near	35.898	[24.260 47.535]	0.270	[0.249 0.291]
	Far	41.265	[25.233 57.297]	0.628	[0.604 0.653]
Textured	Near	36.699	[26.199 47.199]	0.266	[0.245 0.287]
	Far	45.202	[30.501 59.904]	0.587	[0.560 0.613]
Cluttered	Near	24.365	[14.683 34.046]	0.284	[0.267 0.300]
	Far	30.461	[15.549 45.373]	0.611	[0.590 0.631]

Accurate performance for the near third would have a slope of 0.333; accurate performance for the far third would have a slope of 0.666. Accurate intercepts would have a value of zero

Fig. 8 Overall regression trends for the distances perceived in each experiment. Accurate performance indicated in black. In Experiment 1, values for far thirds indicated by dashed lines and near thirds are solid lines
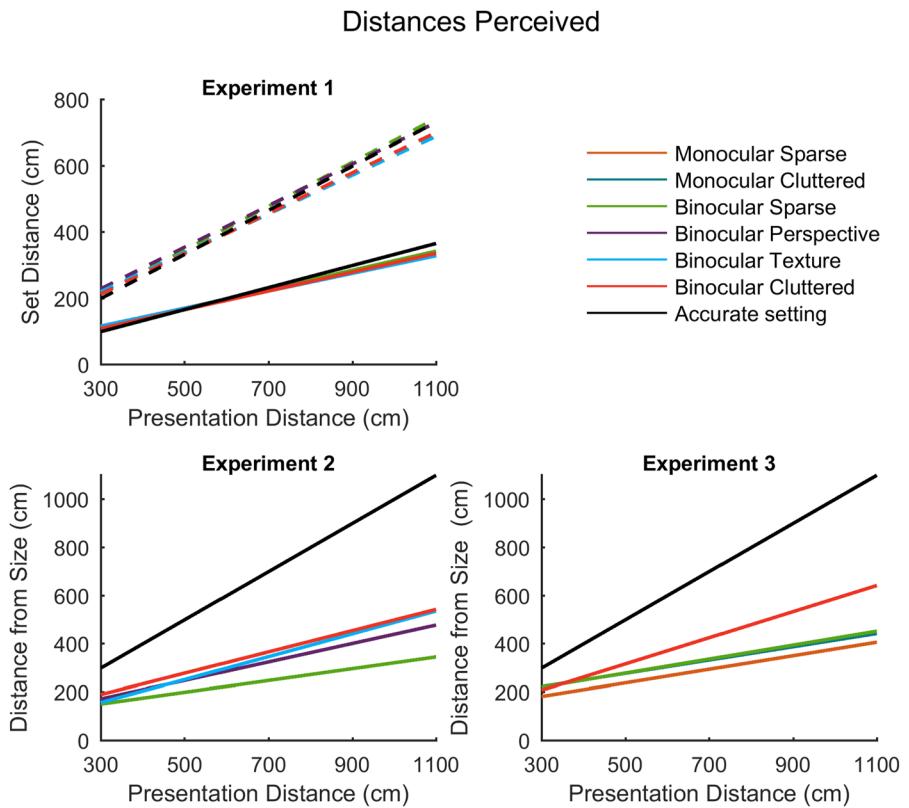

Fig. 9 Overall trends of standard deviation for each environment in all experiments. In Experiment 1, data for near thirds indicated by solid lines and far thirds are dashed lines
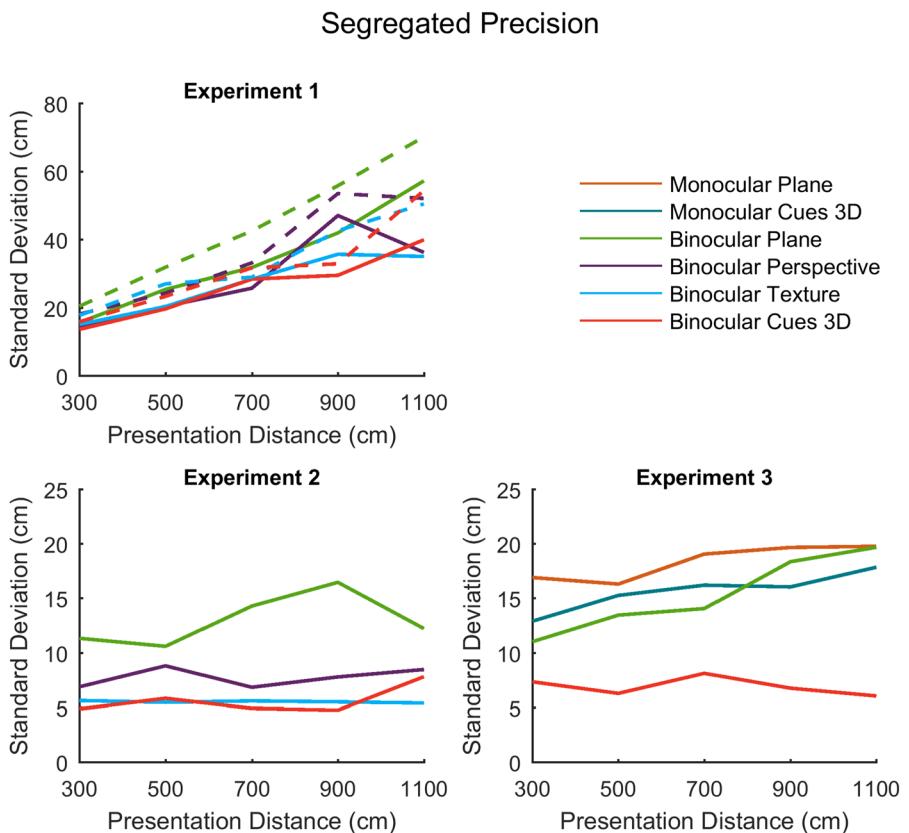
Table 2 Regression analyses of the distance perceived in each environment of Experiment 2, from Eq. 4

Environment	Intercept estimate	Intercept CLs	Slope estimate	Slope CLs
Sparse	78.998	[19.030 138.965]	0.243	[0.149 0.337]
Perspective	58.379	[33.428 83.331]	0.382	[0.297 0.466]
Textured	16.414	[− 9.071 41.899]	0.473	[0.395 0.551]
Cluttered	58.881	[26.885 90.877]	0.440	[0.373 0.508]

Accurate performance would have an intercept of zero and a slope of one

Table 3 Regression analyses of the distance perceived in each environment of Experiment 3 using Eq. 4

Environment	Intercept estimate	Intercept CLs	Slope estimate	Slope CLs
Mono sparse	99.079	[27.181 170.978]	0.280	[0.147 0.414]
Mono cluttered	140.989	[69.611 212.368]	0.275	[0.199 0.351]
Bino sparse	134.775	[6.613 262.938]	0.289	[0.148 0.431]
Bino cluttered	46.370	[− 11.573 104.314]	0.542	[0.449 0.635]

Accurate performance would have an intercept of zero and a slope of one

potential for viewing discomfort that this creates (Banks et al. 2012; Shibata et al. 2011; O'hare et al. 2013).

Compliance with ethical standards

Conflict of Interest The authors declare that they have no conflict of interest

Data availability Available at: http://researchdata.essex.ac.uk/128/.

References

Banks MS, Read JC, Allison RS, Watt SJ (2012) Stereoscopy and the human visual system. SMPTE Motion Imaging J 121(4):24–43

Barfield W, Rosenberg C (1995) Judgments of azimuth and elevation as a function of monoscopic and binocular depth cues using a perspective display. Hum Factors 37(1):173–181

Bingham GP, Bradley A, Bailey M, Vinner R (2001) Accommodation, occlusion, and disparity matching are used to guide reaching: a comparison of actual versus virtual environments. J Exp Psychol Hum Percept Perform 27(6):1314

Bodenheimer B, Meng J, Wu H, Narasimham G, Rump B, McNamara TP, et al. (2007) Distance estimation in virtual and real environments using bisection. In: Proceedings of the 4th symposium on applied perception in graphics and visualization. ACM, pp 35–4

Brenner E, van Damme WJ (1999) Perceived distance, shape and size. Vision Res 39(5):975–986

Carlson V (1960) Overestimation in size-constancy judgments. Am J Psychol 73(2):199–213

Creem-Regehr SH, Willemsen P, Gooch AA, Thompson WB (2005) The inuence of restricted viewing conditions on egocentric distance perception: implications for real and virtual indoor environments. Perception 34(2):191–204

Crompton A, Brown F (2006) Distance estimation in a small-scale environment. Environ Behav 38(5):656–666

Dodgson NA (2004) Variation and extrema of human interpupillary distance. In: Stereoscopic displays and virtual reality systems xi, vol 5291. International Society for Optics and Photonics, pp 36–47

Drascic D (1991) Skill acquisition and task performance in teleoperation using monoscopic and stereoscopic video remote viewing. In Proceedings of the human factors society annual meeting, vol 35, 19. SAGE Publications, Los Angeles, pp 1367–1371

Drascic D, Milgram P (1996) Perceptual issues in augmented reality. In: Stereoscopic displays and virtual reality systems iii, vol 2653. International Society for Optics and Photonics, pp 123–134

Eggleston RG, Janson WP, Aldrich KA (1996) Virtual reality system effects on size-distance judgements in a virtual environment. In: Proceedings of the ieee 1996 virtual reality annual international symposium. IEEE, pp 139–146

Ellis SR, Menges BM (1998) Localization of virtual objects in the near visual field. Hum Factors 40(3):415–431

Glennerster A, Tcheang L, Gilson SJ, Fitzgibbon AW, Parker AJ (2006) Humans ignore motion and stereo cues in favor of a fictional stable world. Curr Biol 16(4):428–432

Hawkins DM, Wixley R (1986) A note on the transformation of chis-quared variables to normality. Am Stat 40(4):296–298

Hibbard P, Goutcher R, Khan N, Hornsey RL (2017a) Manipulations of local, but not global, luminance gradients affect judgements of depth magnitude. J Vis 17(10):1045–1045

Hibbard P, Hornsey RL, Khan NZ (2017b) The contribution of binocular and motion cues to depth quality in complex naturalistic scenes. In: Perception, vol 46, 10. SAGE Publications, London, pp 1219–1220

Hornsey R, Hibbard P (2018) Shape and size constancy in consumer virtual reality. J Vis 18(10):515–515

Hornsey R, Hibbard P (2019) Effects of environmental cues on distance perception in consumer virtual reality. In: Perception, vol 48, 3. SAGE Publications, London, pp 267–267

Hornsey RL, Hibbard PB, Scarfe P (2015) Ordinal judgments of depth in monocularly- and stereoscopically-viewed photographs of complex natural scenes. In: 2015 International Conference on 3D Imaging (IC3D), Liege, Belgium, 2015, pp 1–5. https://doi.org/10.1109/IC3D.2015.7391812

Hornsey R, Hibbard P, Hunter D (2017) Distance perception in consumer virtual reality. J Visi 17(10):1047–1047

Hornsey RL, Hibbard PB, Scarfe P (2020) Size and shape constancy in consumer virtual reality. Behav Res Methods 52:1587–1598

Johnston EB (1991) Systematic distortions of shape from stereopsis. Vis Res 31(7–8):1351–1360

Kim WS, Ellis SR, Tyler ME, Hannaford B, Stark LW (1987) Quantitative evaluation of perspective and stereoscopic displays in three-axis manual tracking tasks. IEEE Trans Syst Man Cybern 17(1):61–72

Kline PB, Witmer BG (1996) Distance perception in virtual environments: effects of field of view and surface texture at near distances. In: Proceedings of the human factors and ergonomics society annual meeting, vol 40, 22. SAGE Publications, Los Angeles, pp 1112–1116

Kopiske KK, Bozzacchi C, Volcic R, Domini F (2019) Multiple distance cues do not prevent systematic biases in reach to grasp movements. Psychol Res 83(1):147–158

Lampton DR, McDonald DP, Singer M, Bliss JP (1995) Distance estimation in virtual environments. In: Proceedings of the human factors and ergonomics society annual meeting, vol 39, 20. SAGE Publications, Los Angeles, pp 1268–1272

Landy MS, Kording K, Trommershauser J (2011) Sensory cue integration. Oxford University Press, Oxford

Lappin JS, Shelton AL, Rieser JJ (2006) Environmental context in uences visually perceived distance. Percept Psychophys 68(4):571–581

Livingston MA, Ai Z, Swan JE, Smallman HS (2009) Indoor vs. outdoor depth perception for mobile augmented reality. In: 2009 ieee virtual reality conference. IEEE, pp 55–62

Loomis JM, Knapp JM et al (2003) Visual perception of egocentric distance in real and virtual environments. Virtual Adapt Environ 11:21–46

Mon-Williams M, Tresilian JR (1999) Some recent studies on the extraretinal contribution to distance perception. Perception 28(2):167–181

Mon-Williams M, Tresilian JR (2000) Ordinal depth information from accommodation? Ergonomics 43(3):391–404

Murgia A, Sharkey PM et al (2009) Estimation of distances in virtual environments using size constancy. Int J Virtual Real 8(1):67–74

Murray RF, Morgenstern Y (2010) Cue combination on the circle and the sphere. J Vis 10(11):15–15

O'hare L, Zhang T, Nefs HT, Hibbard PB (2013) Visual discomfort and depth-of-field. i-Perception 4:156–169

Proffitt DR, Caudek C (2003) Depth perception and the perception of events. In: Handbook of psychology, pp 213–236

Renner RS, Velichkovsky BM, Helmert JR (2013) The perception of egocentric distances in virtual environments—a review. ACM Comput Surv (CSUR) 46(2):23

Rieser JJ, Ashmead DH, Talor CR, Youngquist GA (1990) Visual perception and the guidance of locomotion without vision to previously seen targets. Perception 19(5):675–689

Roumes C, Meehan JW, Plantier J, Menu J-P (2001) Distance estimation in a 3-d imaging display. Int J Aviation Psychol 11(4):381–396

Scarfe P, Hibbard P (2006) Disparity-defined objects moving in depth do not elicit three-dimensional shape constancy. Vis Res 46(10):1599–1610

Scarfe P, Hibbard PB (2011) Statistically optimal integration of biased sensory estimates. J Vis 11(7):12–12

Sedgwick HA (1986) Space perception. Sensory proces and perception

Shibata T, Kim J, Hoffman DM, Banks MS (2011) The zone of comfort: predicting visual discomfort with stereo displays. J Vis 11(8):11–11

Sinai MJ, Krebs WK, Darken RP, Rowland J, McCarley J (1999) Egocentric distance perception in a virutal environment using a perceptual matching task. In: Proceedings of the human factors and ergonomics society annual meeting, vol 43, 22. SAGE Publications, Los Angeles, pp 1256–1260

Singh G, Swan II JE, Jones JA, Ellis SR (2010) Depth judgment measures and occluding surfaces in near-field augmented reality. In: Proceedings of the 7th symposium on applied perception in graphics and visualization. ACM, pp 149–156

Surdick RT, Davis ET, King RA, Hodges LF (1997) The perception of distance in simulated visual displays: a comparison of the effectiveness and accuracy of multiple depth cues across viewing distances. Presence Teleoper Virtual Environ 6(5):513–531

Thomas G, Goldberg JH, Cannon DJ, Hillis SL (2002) Surface textures improve the robustness of stereoscopic depth cues. Hum Factors 44(1):157–170

Todd JT, Christensen JC, Guckes KM (2010) Are discrimination thresholds a valid measure of variance for judgments of slant from texture? J Vis 10(2):20–20

Wickens CD (1990) Three-dimensional stereoscopic display implementation: guidelines derived from human visual capabilities. In: Stereoscopic displays and applications, vol 1256. International Society for Optics and Photonics, pp 2–11

Witmer BG, Kline PB (1998) Judging perceived and traversed distance in virtual environments. Presence 7(2):144–167

Wu B, Ooi TL, He ZJ (2004) Perceiving distance accurately by a directional process of integrating ground information. Nature 428(6978):73–77