Supplemental Material: Perception of Perspective Distortions in Image-Based Rendering

Peter Vangorp [1, 2, 3] Christian Richardt [1] Emily A. Cooper [4] Gaurav Chaurasia [1]
Martin S. Banks [4] George Drettakis [1]

[1] REVES/INRIA Sophia-Antipolis [2] University of Giessen [3] MPI Informatik [4] University of California, Berkeley

Abstract

This document contains supplemental information and clarifications for the paper *Perception of Perspective Distortions in Image-Based Rendering* by Vangorp et al. [2013], published in ACM Transactions on Graphics 32, 4 (Proceedings of ACM SIGGRAPH 2013).

1 Extended Retinal Hypothesis

1.1 Frontoparallel Capture

The derivation of the extended retinal hypothesis assumes frontoparallel capture for simplicity. If the capture was not frontoparallel, an additional coordinate system transformation would be required because the façade features would no longer be aligned to the capture camera axes. However, the derivation for the general case would lead to identical final equations. Thus, the viewing direction of the capture camera does not affect the predictive model.

This can be seen intuitively by considering that we first capture an image and project it back onto the proxy geometry using standard perspective projection. However, this back-projection also works for capture cameras that are not aligned to the façade, or even using cylindrical or spherical projections for 360° panoramas (for which "frontoparallel" is not even defined). The remainder of the derivation only requires that the center of projection of the capture camera be defined.

Eccentricity should be measured from the point on the façade that is closest to the capture camera's center of projection. For a frontoparallel capture camera, that point happens to be at the center of the captured image.

1.2 Simulation Camera Derivation

The following transformation changes from capture to simulation camera coordinates:

$$x_s = (x_c - c \cdot \tan\theta_e) \cdot \cos\theta_s + (z_c - c) \cdot \sin\theta_s \quad (1)$$
$$y_s = y_c \quad (2)$$
$$z_s = (-x_c + c \cdot \tan\theta_e) \cdot \sin\theta_s + (z_c - c) \cdot \cos\theta_s + d \quad (3)$$

Using this transformation, and Equation 6 from the main document, we obtain the projected camera coordinates of the vanishing point for the front face:

$$x_s = -\operatorname{sign}\theta_s \cdot \lim_{t\to\infty} t \cdot \cos\theta_s + [-c \cdot \tan\theta_e \cos\theta_s] \quad (4)$$
$$y_s = 0 \quad (5)$$
$$z_s = \operatorname{sign}\theta_s \cdot \lim_{t\to\infty} t \cdot \sin\theta_s + [c \cdot \tan\theta_e \sin\theta_s] + d \quad (6)$$

The terms in square brackets can be dropped since they are negligible as t approaches infinity if the common trigonometric factor is nonzero, or they are zero if the common trigonometric factor is zero.

And the projected camera coordinates of the vanishing point for the side face:

$$x_s = -c \cdot \tan\theta_e \cdot \cos\theta_s \quad (7)$$
$$y_s = 0 \quad (8)$$
$$z_s = c \cdot \tan\theta_e \cdot \sin\theta_s + d \quad (9)$$

We then perform perspective projection with focal length f_s resulting in Equations 7–10 in the main document.

2 Experimental Design

2.1 Stimulus Generation

The stimulus images were created using the PBRT offline renderer [Pharr and Humphreys 2010] in two passes:

1. The capture camera created a wide-angle frontoparallel image of the three different realistically-dimensioned façade designs from a distance of 40 m. The capture camera had a field of view of 107°×20° and a very high resolution of 17200×2224 pixels to avoid noticeable aliasing artifacts in the next pass.
2. The captured image was projectively textured onto a single plane with a slant angle θ_s of 0°, ±15° or ±30° with respect to the optical axis of the simulation camera. The simulation camera was pointed straight at a corner at a distance of 40 m and had a field of view of 21°×16° and a resolution of 2048×1536 pixels. This resolution was downsampled for display on different devices.

Both renderpasses used high-quality antialiasing and EWA texture filtering [Greene and Heckbert 1986] to ensure artifact-free final images.

2.2 Display Setup

We wrote our experiments in MATLAB, using the Psychophysics Toolbox [Kleiner et al. 2007] to display stimuli and collect keyboard input. The hinge device used a potentiometer to record the angle between the sides. The TV can be treated like a normal PC monitor

Display	Diag.	Image diag.	View. distance	retinal factor
Phone	3.5"	67.4 mm	341 mm	2.37
Tablet	9.7"	207 mm	522 mm	1.18
PC	24"	458 mm	848 mm	0.87
TV	55"	920 mm	1450 mm	0.74

Table 1: *The four display conditions used in our experiments. Columns are: the screen diagonal, the diagonal of the stimulus image, the corresponding preferred viewing distance [Cooper et al. 2012] and retinal minification factor $v/(M \cdot f_s)$.*

as it can be connected directly to a computer. For the phone and the tablet, we instead use VNC to relay the display content from the computer over a wireless network. On these devices, we also use a bluetooth keyboard for input.

2.3 Screening of Participants

The hinge angle-matching task had precise instructions on how to align the hinge device with the façade and reproduce the angle. The rating task requires participants to use the same internal decision-making process to provide self-consistent ratings. Both tasks involve response mapping between the observed angle in the displayed stimulus and either the observed angle of the physical hinge device or the 5-point rating. Some participants are known to have difficulty performing such response mapping tasks [Watt et al. 2005]. Combined with the large number of conditions in this within-subject design, performing both tasks for extended periods requires significant effort and motivation from the participants. It is accepted practice to verify in advance that participants will be able to perform such an extensive experiment, without biasing the result. Therefore participants were screened using short versions of the experiments (112 trials). We excluded 3 participants: one for not following instructions precisely on setting the hinge, one for rating the same stimulus at two extremes of the scale 3 times each, and one who had multiple inconsistent responses to the same hinge stimulus.

3 Experiment 1: Hinge Angle Matching

3.1 Additional Results Figures

Figure 1 shows the results of the hinge angle-matching experiment for each of the 4 display devices. There are no significant differences visible.

Figure 2 shows the results of the hinge angle-matching experiment for the 6 participants separately. There are differences between participants, but the repeated measures design allows us to study the effects of interest while keeping the variability due to individual differences low.

Many more plots are available on the project web page.

3.2 Analysis of Variance

Table 2 shows the ANOVA table for the factors display device, simulation angle, eccentricity angle, façade depth and façade design. Compared to the ANOVA in the main document that omitted the façade design factor, all significant effects remain significant but with a potentially reduced effect size. The main effect of façade depth and the interaction between simulation and eccentricity angle are now reduced to medium-large effect sizes ($0.13 < \eta^2_G < 0.26$). There are no additional statistically significant effects of façade design with appreciable effect sizes.

4 Experiment 2: Angle Rating

4.1 Additional Results Figures

Figure 3 shows the results of the angle-rating experiment for each of the 4 display devices. There are no significant differences visible.

Figure 4 shows the results of the angle-rating experiment for the 6 participants separately. There are differences between participants as can be expected in rating tasks, but the pattern that stands out across all participants is the blue diagonal of angles rated close to a right angle.

Source	df	SS	MS	F	p	Sig	η^2_G
Display Device (Dev)	3	10716	3572	7.939	0.0021	⋆	0.051
Simulation Angle (Sim)	4	2333	583.2	7.102	0.000995	⋆	0.011
Eccentricity Angle (Ecc)	3	168059	56020	50.05	4.78e-08	⋆	0.456
Façade Depth (Dep)	2	56364	28182	61.84	2.34e-06	⋆	0.219
Façade Design (Faç)	2	2647	1324	0.773	0.487		0.013
Dev×**Sim**	12	168.9	14.07	0.551	0.872		0.001
Dev×**Ecc**	9	1387	154.2	1.362	0.234		0.007
Dev×**Dep**	6	626.9	104.49	1.647	0.169		0.003
Dev×**Faç**	6	1655	275.9	2.038	0.0912		0.008
Sim×**Ecc**	12	52016	4335	11.39	1.92e-11	⋆	0.206
Sim×**Dep**	8	332.8	41.60	1.228	0.308		0.002
Sim×**Faç**	8	699.1	87.39	2.067	0.0626		0.003
Ecc×**Dep**	6	7783	1297	9.135	1.03e-05	⋆	0.037
Ecc×**Faç**	6	9315	1552.5	4.176	0.00362	⋆	0.044
Dep×**Faç**	4	6121	1530.2	8.003	0.000506	⋆	0.030
Dev×**Sim**×**Ecc**	36	2050	56.93	2.142	0.000592	⋆	0.010
Dev×**Sim**×**Dep**	24	524.2	21.84	0.884	0.622		0.003
Dev×**Sim**×**Faç**	24	629.1	26.21	1.125	0.328		0.003
Dev×**Ecc**×**Dep**	18	1285	71.39	1.933	0.0225	⋆	0.006
Dev×**Ecc**×**Faç**	18	850	47.24	1.163	0.309		0.004
Dev×**Dep**×**Faç**	12	232.3	19.36	0.662	0.781		0.001
Sim×**Ecc**×**Dep**	24	1308	54.50	1.745	0.0268	⋆	0.006
Sim×**Ecc**×**Faç**	24	6663	277.61	2.957	5.29e-05	⋆	0.032
Sim×**Dep**×**Faç**	16	455.2	28.45	1.022	0.443		0.002
Ecc×**Dep**×**Faç**	12	4432	369.3	5.123	8.19e-06	⋆	0.022
Dev×**Sim**×**Ecc**×**Dep**	72	2354	32.70	1.224	0.121		0.012
Dev×**Sim**×**Ecc**×**Faç**	72	2125	29.52	1.096	0.292		0.010
Dev×**Sim**×**Dep**×**Faç**	48	1291	26.89	1.088	0.333		0.006
Dev×**Ecc**×**Dep**×**Faç**	36	1502	41.71	1.549	0.0336	⋆	0.007
Sim×**Ecc**×**Dep**×**Faç**	48	1906	39.71	1.382	0.0613		0.009
Dev×**Sim**×**Ecc**×**Dep**×**Faç**	144	3236	22.48	0.884	0.818		0.016

Table 2: *Results of our repeated-measures ANOVA on the hinge data. The columns list: sources of variance, their degrees of freedom (df), sum of squares (SS), mean square (MS), F-statistic, p-value, significance code for $\alpha = 0.05$, and generalized η^2 effect size.*

Many more plots are available on the project web page.

4.2 Follow-up Experiment

Figure 5 shows the results of the follow-up experiment with real stimuli for both the hinge angle-matching task (a) and the rating task (b). The results are qualitatively similar to the results of the main experiments with synthetic stimuli, but overall larger ranges of angles and ratings were obtained.

5 A Predictive Model for Perspective Distortion in Street-level IBR

5.1 Flattening of Perceived Angles

The retinal hypothesis (Equations 12–16 in the main document) predicts different results for the different devices because the viewing distances relative to the COP differs across devices. However, Figure 1 did not reveal systematic differences across devices. Our explanation for the lack of a device effect is that the effect of distance from the COP is overshadowed by the overall compression of responses towards 90°, likely due to the familiarity with 90° balcony shapes such as the ones in our stimuli [Yang and Kubovy 1999, Perkins 1972]. For this reason, we assumed in Figure 7(a) in the main document a modified viewing distance $v' = 1$, i.e., the viewer is at the COP.

Additional to compensation of viewing distance to the COP distance, an overall flattening effect independent of the stimulus parameters can occur because binocular viewing of flat, non-stereoscopic display surfaces provides binocular depth cues in conflict with the pictorial cues from converging parallel lines [Yang and Kubovy 1999]. This

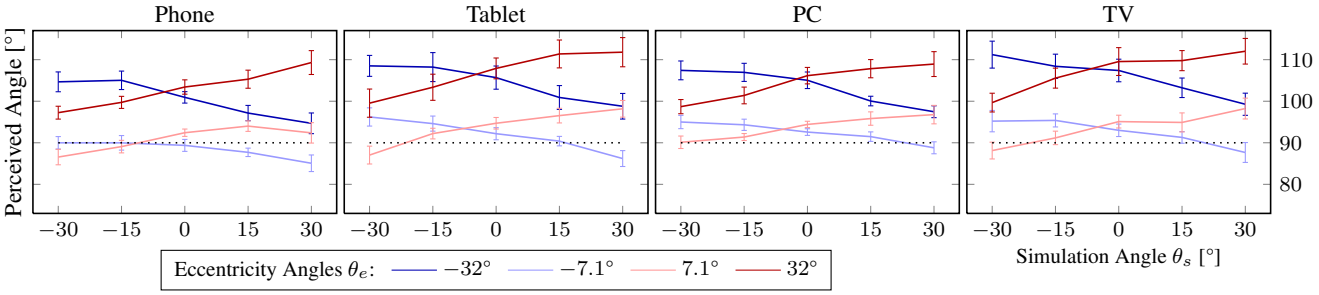

Figure 1: *Hinge angle-matching results for different display devices. Each line corresponds to an eccentricity angle; error bars indicate one standard error above and below the mean. The horizontal dotted line at 90° represents the scene hypothesis.*

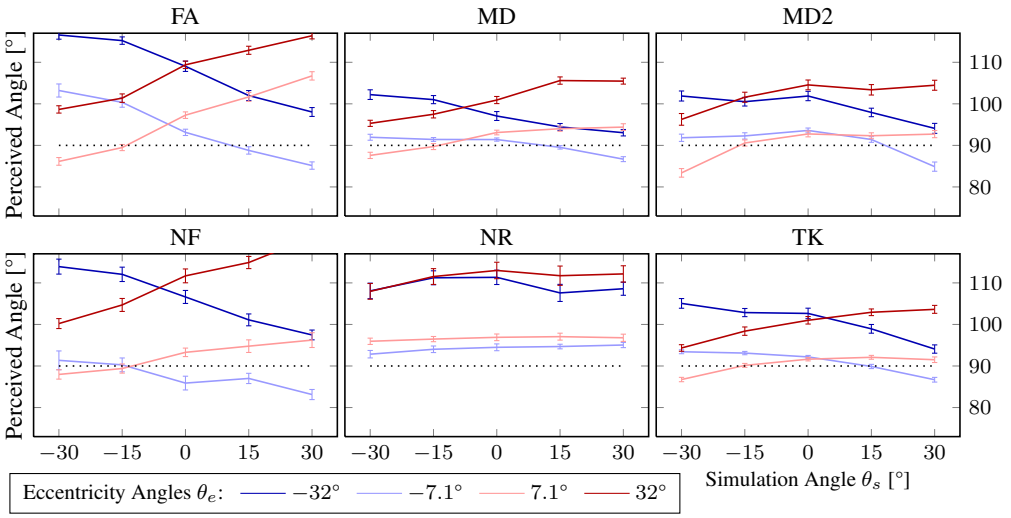

Figure 2: *Hinge angle-matching results for different participants. Each line corresponds to an eccentricity angle; error bars indicate one standard error above and below the mean. The horizontal dotted line at 90° represents the scene hypothesis.*

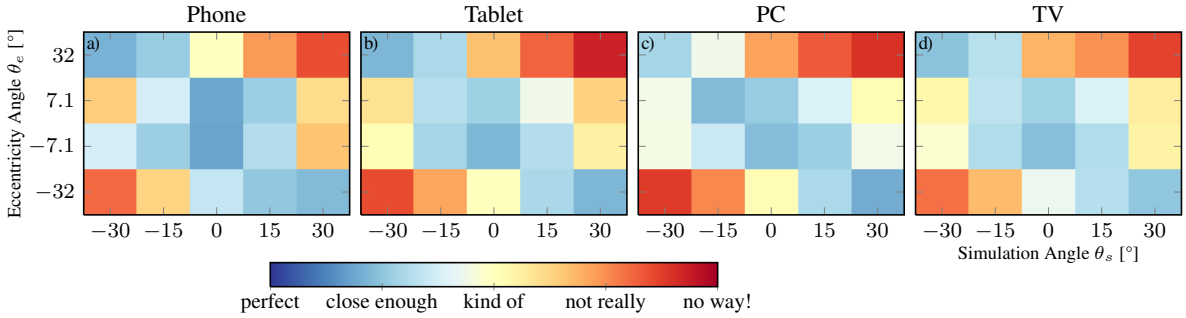

Figure 3: *Interpolated medians of rating results for different display devices.*

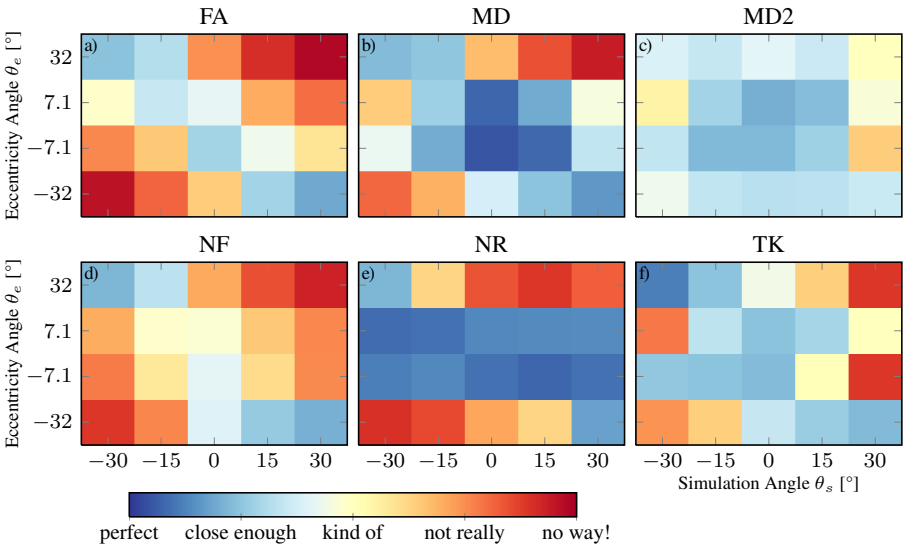

Figure 4: *Interpolated medians of rating results for different participants.*

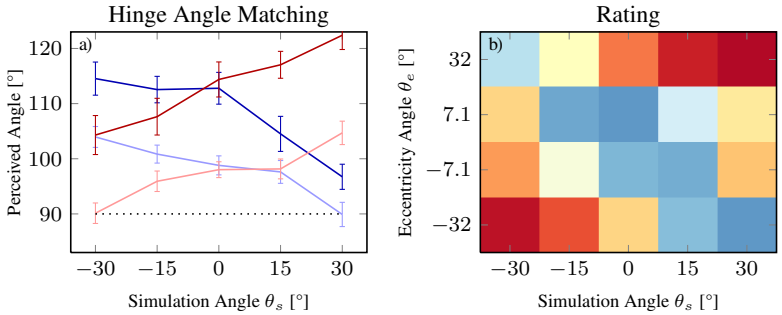

Figure 5: *Results of the follow-up experiment with real stimulus images. (a) Hinge angle-matching results. (b) Rating results.*

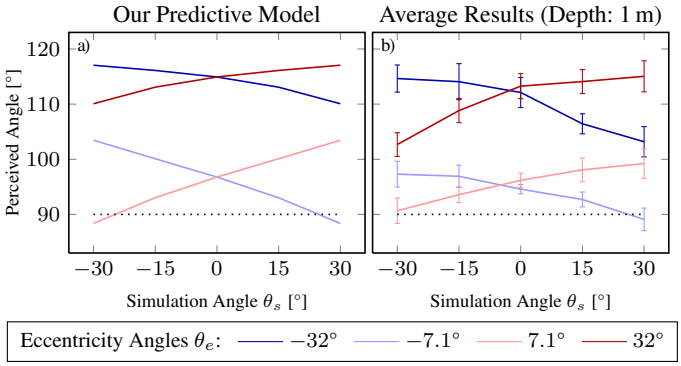

Figure 6: *Our predictive model including the flattening effect (a) compared to the hinge angle-matching results (b). Compare to Figure 10 in the main document to see that these predicted perceived angles match the data slightly better.*

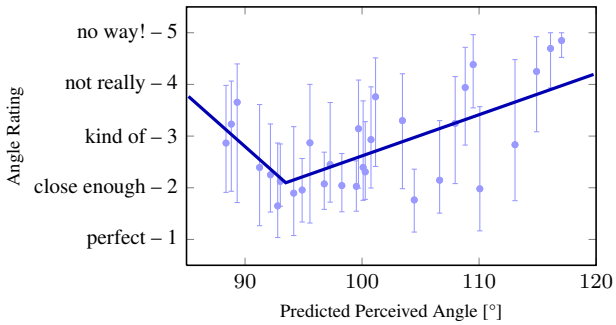

Figure 7: *The best piecewise-linear fit from perceived angles, as predicted by our model including the flattening effect, to the ratings in our experiment. Compare to Figure 11 in the main document to see that the effect of flattening significantly affects this fit.*

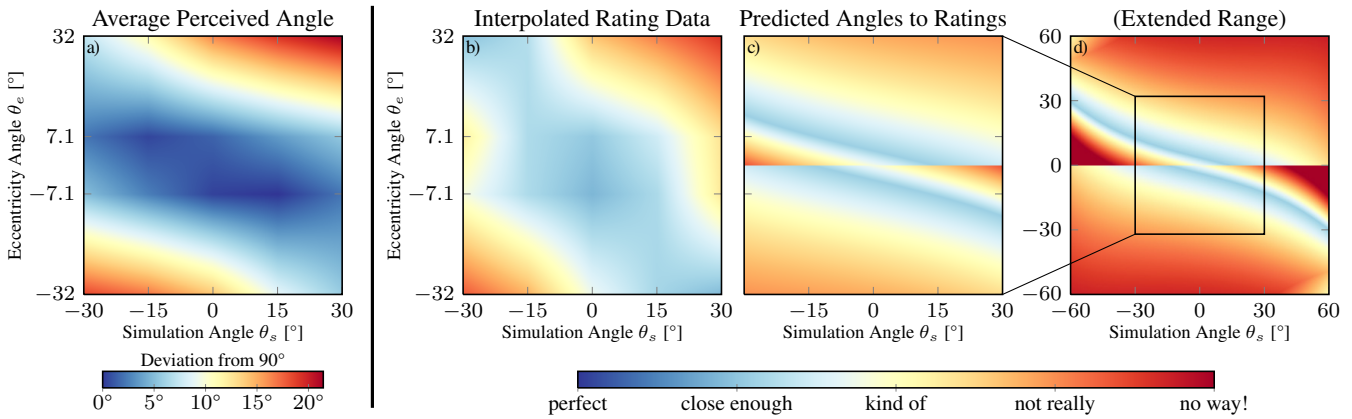

Figure 8: *(a) The deviation of average perceived angles (experiment 1) from 90° has a similar structure to (b) the interpolated medians of ratings from experiment 2 (both interpolated). (c) Perceived angles predicted by the predictive model including the flattening effect, and then mapped to ratings using Figure 7, and (d) the same for an extended domain. Compare to Figure 12 in the main document to see that the flattening has an adverse effect on these rating predictions.*

flattening causes the scene as a whole to appear flatter, and thus the observed angles appear more obtuse. Notice that this effect partially counteracts the regression towards the scene hypothesis for obtuse perceived angles, but reinforces it for acute perceived angles.

The effect of flattening can be modeled as a modified viewing distance $v'' < v'$, obtained by fitting the retinal minification factor. The optimal retinal minification factor, fitted across all devices, was 0.503. In other words, the effect of flattening is as if participants viewed the stimuli from a modified viewing distance v'' of approximately half the COP distance. This also leads to different fits for depth falloff function and the piecewise-linear mapping from perceived angles to ratings (Figure 7).

Figure 6 demonstrates that this additional parameter slightly improves the fit of the predictive model for perceived angles, but Figure 8 shows that it has an adverse effect on the accuracy of predicted ratings. This might be due to the differences in the tasks: in the hinge angle-matching task participants compare an image of an angle to a physical angle of the hinge device, while in the rating task participants compare the same image of an angle to the concept of a right angle.

6 Guidelines for Street-Level IBR

6.1 Reparameterization

For the practical guidelines for street-level IBR applications, we first need to convert a given novel simulation camera position, $c_p = (x, y, z)$, specified in the capture camera coordinate system of an image captured at a distance c from the façade, and simulation camera angle θ, to the parameters θ_s, θ_e and d used in our predictive model:

$$\theta_s = -\theta \tag{10}$$

$$\theta_e = \arctan \frac{x + (c - z) \cdot \tan\theta}{c} \tag{11}$$

$$d = (c - z)/\cos\theta \tag{12}$$

6.2 Guidelines for Display and Capture

After the above reparameterization, the predicted rating of any corner from any simulation camera position and orientation can be calculated using the following equations in the main document. The perceived angle predicted by the extended retinal hypothesis can be calculated by evaluating Equations 13, 15 and 16. The predictive model then produces the perceived angle using Equation 18 with the fitted parameters $depth_0 = 0.50\,\mathrm{m}$, $depth_1 = 0.81\,\mathrm{m}$, $y_0 = 0.23$, and $y_1 = 0.48$, in Equation 19. Finally, the predicted rating is calculated using Equation 20.

References

COOPER, E. A., PIAZZA, E. A., AND BANKS, M. S. 2012. The perceptual basis of common photographic practice. *Journal of Vision 12*, 5, 8:1–14.

GREENE, N., AND HECKBERT, P. S. 1986. Creating raster omnimax images from multiple perspective views using the elliptical weighted average filter. *IEEE Computer Graphics & Applications 6*, 6, 21–27.

KLEINER, M., BRAINARD, D. H., AND PELLI, D. G. 2007. What's new in Psychtoolbox-3? *Perception 36*, 14. ECVP Supplement.

PERKINS, D. N. 1972. Visual discrimination between rectangular and nonrectangular parallelopipeds. *Attention, Perception, & Psychophysics 12*, 5, 396–400.

PHARR, M., AND HUMPHREYS, G. 2010. *Physically Based Rendering: From Theory to Implementation*, 2nd ed. Morgan Kaufmann.

VANGORP, P., RICHARDT, C., COOPER, E. A., CHAURASIA, G., BANKS, M. S., AND DRETTAKIS, G. 2013. Perception of perspective distortions in image-based rendering. *ACM Transactions on Graphics 32*, 4. Proceedings of ACM SIGGRAPH 2013.

WATT, S. J., AKELEY, K., ERNST, M. O., AND BANKS, M. S. 2005. Focus cues affect perceived depth. *Journal of Vision 5*, 10, 7:834–862.

YANG, T., AND KUBOVY, M. 1999. Weakening the robustness of perspective: Evidence for a modified theory of compensation in picture perception. *Attention, Perception, & Psychophysics 61*, 3, 456–467.