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

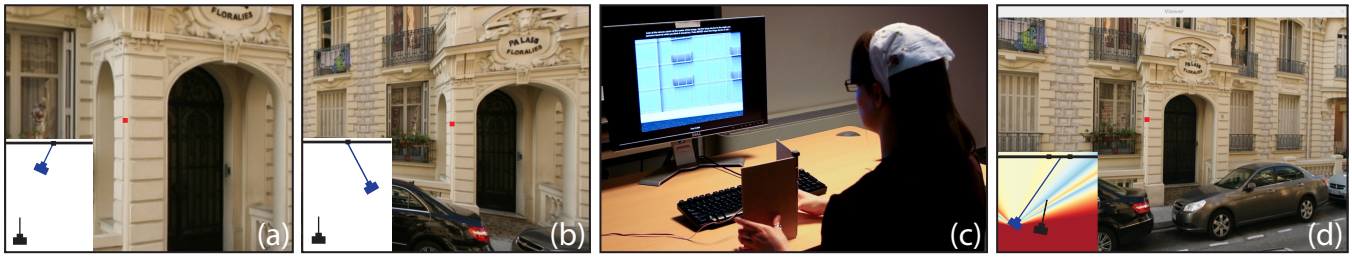

Figure 1: *(a–b) Two images of street-level image-based navigation, using a single captured panorama: both views are away from the original position of the photo. View (a) is most likely perceived as not distorted, while view (b) is perceived to be very distorted. We extended vision science theories on picture perception to predict perceived distortion for such scenes. We designed and ran an experiment (c) to measure perceived distortion. The results are used in an interactive application (d) to predict the quality of images: blue zones in the inset are regions in which the user can navigate without seeing distortions. Capture cameras are represented by black icons, virtual cameras by blue icons.*

Abstract

Image-based rendering (IBR) creates realistic images by enriching simple geometries with photographs, e.g., mapping the photograph of a building façade onto a plane. However, as soon as the viewer moves away from the correct viewpoint, the image in the retina becomes distorted, sometimes leading to gross misperceptions of the original geometry. Two hypotheses from vision science state how viewers perceive such image distortions, one claiming that they can compensate for them (and therefore perceive scene geometry reasonably correctly), and one claiming that they cannot compensate (and therefore can perceive rather significant distortions). We modified the latter hypothesis so that it extends to street-level IBR. We then conducted a rigorous experiment that measured the magnitude of perceptual distortions that occur with IBR for façade viewing. We also conducted a rating experiment that assessed the acceptability of the distortions. The results of the two experiments were consistent with one another. They showed that viewers' percepts are indeed distorted, but not as severely as predicted by the modified vision science hypothesis. From our experimental results, we develop a predictive model of distortion for street-level IBR, which we use to provide guidelines for acceptability of virtual views and for capture camera density. We perform a confirmatory study to validate our predictions, and illustrate their use with an application that guides users in IBR navigation to stay in regions where virtual views yield acceptable perceptual distortions.

CR Categories: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism

Keywords: Image-based rendering, perception, human vision

Links: DL PDF WEB

1 Introduction

Image-based rendering (IBR) provides realistic 3D imagery with a few photographs as input [Shum et al. 2006], thus avoiding the manual and time-consuming content creation pipeline of traditional computer graphics. Recent street-level navigation systems (e.g., Google Maps Street View™ [Vincent 2007], Bing Maps Streetside™, or Mappy Urban Dive™) use a simple form of IBR consisting of a panorama and piecewise planar approximations of the ground and building façades. Despite their simplicity, such systems create reasonably compelling 3D experiences: we will refer to these as *street-level IBR*. However, the resulting images are only correct when viewed from where the original photographs were taken; when moving away from this point, distortions can become quite large, as shown in Fig. 1(b). Because of this, such systems typically restrict viewers to be near one of the capture points.

Two terms are important for understanding the images created in street-level IBR and users' perceptions of those images. *Image distortion* refers to retinal images (in picture viewing) that are not the same as the images created when viewing the original 3D scenes. Such distortions can occur because the displayed image is distorted and/or because the viewer is not positioned at the center of projection (COP). *Perceptual outcome* refers to the viewer's perception derived from the retinal image.

A key insight in our work is to make the link between distortions in street-level IBR and what studies of human vision tell us about the resulting perceptual outcomes. The vision science literature provides two useful hypotheses concerning the perception of pictures: the *scene hypothesis* and the *retinal hypothesis*. The scene hypothesis states that viewers compensate for incorrect viewing position, so the perceptual outcome is much closer to the original 3D scene than dictated by the distorted retinal image. To understand this, note that the viewer must be positioned at the picture's center of projection for the retinal image to be a faithful copy of the image that would be created by viewing the original 3D scene. The retinal hypothesis, on the other hand, states that viewers do not compensate for incorrect position; rather the perceptual outcome is dictated by the distorted

ACM Reference Format
Vangorp, P., Richardt, C., Cooper, E., Chaurasia, G., Banks, M., Drettakis, G. 2013. Perception of Perspective Distortions in Image-Based Rendering. ACM Trans. Graph. 32, 4, Article 58 (July 2013), 11 pages. DOI = 10.1145/2461912.2461971 http://doi.acm.org/10.1145/2461912.2461971.

DOI: http://doi.acm.org/10.1145/2461912.2461971

retinal image. Vision science literature, reviewed in Sec. 2, shows that each hypothesis is valid, but for different contexts.

The situation is more complex in street-level IBR. In these systems, an image is first captured by photographing a façade from a given position. This image is then projected onto simplified geometry, like a 3D polygon. The 3D polygon is in turn viewed from a new camera, and projected into 2D to form the final image. Distortions occur because the new camera has a different COP than the one associated with the original photograph. Such situations have not been studied extensively in vision science, but a similar case arises in pictures within pictures [Pirenne 1970]. When a photograph within the photograph is slanted, there are two COPs, one for the photographed 3D scene (that COP is usually on the central surface normal) and one for the slanted photograph within the scene (that COP is not on the central surface normal). Viewers can compensate for their incorrect position with respect to the first COP, but are generally unable to compensate for incorrect position with respect to the second COP. Thus, compensation occurs for pictures created with one camera and not for content created by a second camera. Street-level IBR creates both of these scenarios. We will use concepts from vision science to understand how image distortions created in street-level IBR systems are perceived and thereby to determine how best to create such imagery without objectionable perceptual distortions.

We make three primary contributions:

- We modify the retinal hypothesis to extend it to street-level IBR. The quantitative model that emerges from this modification uses perspective information to generate predicted 3D percepts.
- In two perceptual experiments, we determine how well the extended retinal and scene hypotheses predict perceptual outcomes. We find that outcomes fall in-between the two predictions and that those outcomes are very consistent with viewers' ratings of how acceptable different views are.
- We develop a predictive model for perceptual distortions in street-level IBR, and use it to present guidelines for acceptability of novel views and capture camera density. We also perform a validation study of our predictions, and illustrate their use to guide users of street-level IBR so that images of novel views are acceptable during navigation.

2 Related Work

Perception of Artifacts in Image-based Rendering Texture mapping has been widely used to represent surface color and also shape details that are too small to model geometrically. An example is projective texture mapping for urban visualizations [Debevec et al. 1998]. In such cases, the shape details do not change when the simulated (or virtual) viewpoint changes as they would when changing viewpoint in the original scene; as a consequence, the image becomes distorted. View-dependent texture mapping [Debevec et al. 1998] and unstructured Lumigraph rendering [Buehler et al. 2001] provide ways to choose among or blend between multiple textures that were captured from different viewpoints. In practice, the texture viewpoints are not sampled densely, so novel simulated viewpoints can still create image distortions – angle distortions and blending artifacts – that are quite objectionable.

Previous work on the perception of image distortions in IBR has focused on the transition artifacts that occur when the simulated viewpoint moves between input viewpoints. Occlusion boundaries, particularly their synchronization with camera motion, are key determinants of the perceived quality of transitions [Morvan and O'Sullivan 2009]. Transitions are more acceptable when exact edge correspondences, coherent motion and homogeneous regions are provided [Stich et al. 2011]. Images for simulated novel viewpoints are more acceptable when the transitions are faster, there are fewer transition artifacts, and the distance to the viewpoint of the used texture is small [Vangorp et al. 2011]. Geometric distortions due to viewing distances inconsistent with the field of view of the image have been studied in the context of traditional rendering rather than IBR, confirming that observers make the most accurate geometric estimates of the scene when they are close to the consistent viewpoint [Steinicke et al. 2011].

Despite these advances in our understanding, previous work has not provided an explanation of why some image distortions are acceptable, while others are not. We claim that studies of human picture perception provide data and concepts that prove to be useful for a better implementation of street-level IBR systems.

Vision Science Literature on Picture Perception Pictures, which are defined as 2D representations of 3D scenes based on perspective projection, can yield compelling impressions of 3D structure. To estimate 3D structure, humans use a variety of depth cues that are potentially available in pictures: cues based on perspective, lighting, atmospheric effects, and triangulation. Here we focus on perspective-based cues because they are most affected by the image distortions that occur in street-level IBR.

Perspective-based cues can be used to determine the 3D layout of a scene up to a scale factor. For instance, when a photograph is taken of a slanted rectangle, the objectively parallel edges of the rectangle are imaged as non-parallel. If one extends the lines until they intersect at the vanishing point, the angle between a line from the viewer to the vanishing point and a line from the viewer to the rectangle specifies the slant and tilt of the rectangle [Sedgwick 1991]. When a perspectively correct picture is viewed from the COP, people are quite accurate at recovering the 3D geometry of the original scene, including the slants of surfaces in that scene [Smith and Smith 1961, Cooper et al. 2012]. If the viewer's eye is offset from the COP, perspective-based cues no longer specify the original 3D scene; instead, they specify a different, distorted scene. Research in picture perception has been focused on how those distortions affect the perception of 3D shape, and puts forward the two hypotheses defined in Sec. 1: the *scene* and *retinal* hypotheses. In some situations, the experimental evidence favors the scene hypothesis. For example, when viewers are left or right of the COP and view the picture and its frame with both eyes, they compensate for their incorrect viewing position and perceive 3D structure reasonably accurately [Rosinski et al. 1980, Vishwanath et al. 2005]. Some have claimed that this compensation is based on the use of familiar shapes, such as cubes [Perkins 1972, Yang and Kubovy 1999] while others have claimed that it is based on measurement of the orientation of the surface of the picture [Wallach and Marshall 1986, Vishwanath et al. 2005]. In other situations, the experimental evidence favors the retinal hypothesis, even for small off-axis displacements [Banks et al. 2009]. When the slant of a pictured object is nearly perpendicular to the picture surface, little compensation for off-axis viewing occurs [Goldstein 1987, Todorović 2008]. When viewers are too close to or too far from the picture, they do not compensate for the induced image distortions and therefore perceive 3D structure incorrectly [Adams 1972, Cooper et al. 2012, Lumsden 1983, Todorović 2009].

Thus the scene and retinal hypotheses account best for perceptual outcomes in different situations. Our goal is to study how well these hypotheses predict perceptual distortions in street-level IBR.

Street-level Image-based Viewing Another key goal is to use our findings to provide practical guidelines in an application setting. We will focus on a simplified image-based setup which is akin to existing visualizations of street-level imagery, e.g., Google Maps Street View™, Bing Maps Streetside™ and Mappy Urban

Dive™. While exact details of these systems are not always available, they use panoramic images captured at discrete points along a path, and rendered using the equivalent of view-dependent texture mapping [Debevec et al. 1998] onto a single planar proxy for each façade, or a similar technique. In most of these systems, transitions between viewpoints occur as a fast blurred mix between the viewpoints, possibly with a blending approach (e.g., akin to Buehler et al. [2001]). Our analysis also applies to other types of street-level IBR such as Microsoft Photosynth™ [Snavely et al. 2006], Street Slide [Kopf et al. 2010], and even multi-perspective images [Yu et al. 2010] as long as corners in façades are only deformed by perspective projections and not by the image deformations applied to align and stitch images into panoramas.

In this paper, we assume façades are captured in a similar manner: using panoramas (or wide-angle images) at discrete points along a path and reprojected onto a planar proxy for the façade and ground.

3 Extended Retinal Hypothesis

In this section, we describe an extended retinal hypothesis for street-level IBR. In describing the hypothesis, we consider the viewing of corners on building façades (e.g., corners of balconies). The result is a prediction for the perceived angle of such corners. The four steps involved in the process are:

1. **Capture:** Street-level panoramas (or photographs) are captured, and camera positions and orientations are registered.
2. **Projection:** The captured imagery is mapped onto simplified geometry, often only a ground plane and a few façade planes.
3. **Simulation:** This scene is visualized from the different viewpoint of a virtual or *simulation* camera.
4. **Display:** The resulting image is presented on a display device and viewed by the user.

As explained in Sec. 2, the retinal hypothesis states that the perceptual outcome is a direct interpretation of the retinal image without compensation for possibly incorrect positions of the simulation camera or viewer. Vanishing points are formed by extensions of edges in the façade and the ground plane, and are used by viewers to estimate the geometry.

To develop our extended retinal hypothesis, we derive the coordinates of vanishing points for an arbitrary corner seen in the simulation camera image (e.g., a balcony corner). This will allow us to predict the angle perceived according to the projected retinal image. We restrict our derivation and experiments to planar proxies and box-like balconies. These are reasonable approximations for many façades. We show in the validation study how our results can be applied in other situations.

As sketched in Fig. 2, the capture camera coordinate system has its origin at the camera's nodal point (optical center), the x-axis points to the right, the y-axis points up and the z-axis points forward. The camera captures a façade with corners (in black). We use the following two angles extensively: the *eccentricity angle*, denoted θ_e, defined by the direction normal to the façade (in grey on the left) and the line between the corner and the origin, and the *simulation angle*, denoted θ_s, which is the angle between the normal direction and the line between the simulation camera nodal point and the corner (right). The left inset shows the image of a capture camera, while the righthand two insets show simulation (virtual) images.

3.1 Capture

Consider a corner on the façade, with front and side faces (blue and red in Fig. 3). The vanishing point of the front face of the corner is a

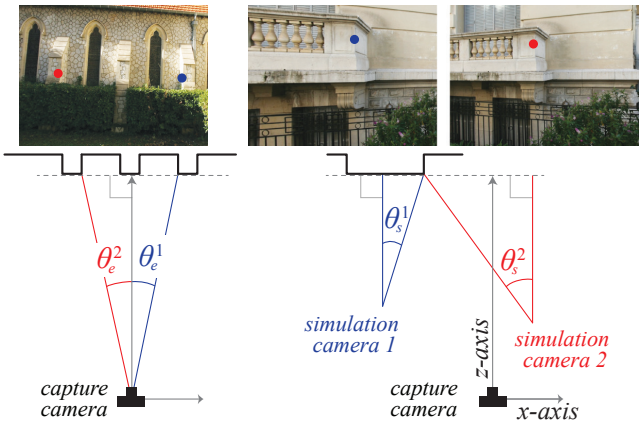

Figure 2: *Left: Top-view of capture camera and two* eccentricity angles, θ_e^1 *and* θ_e^2. *Right: two simulation cameras, their* simulation angles θ_s^1 *and* θ_s^2, *and corresponding virtual images. See how simulation angle affects final image quality. Depending on which corner we look at, eccentricity also affects distortion.*

point with parameter t that runs to infinity on the façade, to the side that the simulation camera will be turned:

$$x_c = -\operatorname{sign}\theta_s \cdot \lim_{t\to\infty} t, \quad y_c = 0, \quad z_c = c, \tag{1}$$

where c is the *capture* z-distance from the corner to origin. We maintain the limit notation to later differentiate cases occurring in the capture and simulation projections.

The vanishing point of the side face of the corner is similarly defined as a point with parameter t that runs to infinity on that side away from the camera:

$$x_c = c \cdot \tan\theta_e, \quad y_c = 0, \quad z_c = \lim_{t\to\infty} t. \tag{2}$$

The image plane of the capture camera is a frontoparallel plane at focal length f_c *in front of* the nodal point. After perspective projection, the vanishing points are on the image plane, at coordinates x_c', y_c':

$$x_c' = f_c \cdot x_c/z_c, \quad y_c' = f_c \cdot y_c/z_c. \tag{3}$$

We now derive the capture-camera image-space coordinates for the vanishing points of the two faces. For the front face:

$$x_c' = -\operatorname{sign}\theta_s \cdot \lim_{t\to\infty} t, \quad y_c' = 0. \tag{4}$$

Note that $x_c' = \pm\infty$ if the façade is frontoparallel to the capture camera. For the side face:

$$x_c' = 0, \quad y_c' = 0. \tag{5}$$

The insets of Fig. 3 (left) illustrate this step.

3.2 Projection

The projection step in the IBR pipeline consists of reprojecting the panorama onto a planar proxy. Assume that the planar proxy is at distance c; to project the image onto it, we replace the capture camera with a projector at the same position, orientation and focal length f_c. The projector coordinate system is the same as the capture camera coordinate system. The effect on vanishing point coordinates is simply to set the coordinate z_c to c. Projected camera coordinates of the vanishing points for the front face are:

$$x_c = -\operatorname{sign}\theta_s \cdot \lim_{t\to\infty} t, \quad y_c = 0, \quad z_c = c \tag{6}$$

and for the side face: $x_c = 0$, $y_c = 0$, $z_c = c$.

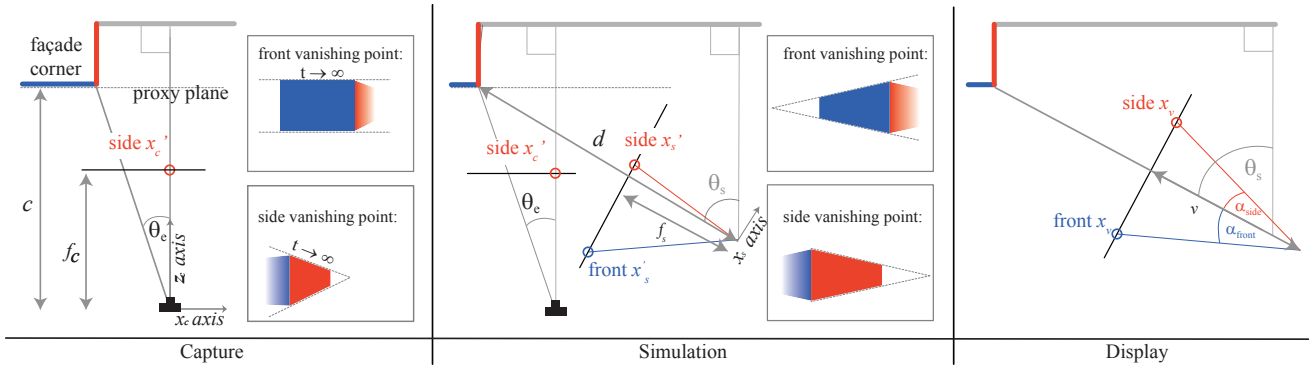

Figure 3: *Left: Top view of capture camera geometry; the insets show the vanishing points (VPs) as seen in the screen for front and side of balcony. Middle: Top view of simulation camera geometry: again insets show vanishing points. Right: Top view of display geometry.*

3.3 Simulation

We assume that the simulation camera points straight at the corner with eccentricity angle θ_e, from a distance d and simulation angle θ_s (Fig. 3, center). We transform the vanishing points to the simulation camera coordinate system and then perform perspective projection of the simulation camera (i.e., the IBR itself), with focal length f_s. This transformation and the coordinates x_s' and y_s' of the vanishing points in simulation camera coordinates after perspective projection are derived in the supplemental material. These coordinates are:

Front:

$$x_s' = f_s \frac{-\operatorname{sign}\theta_s \cdot \lim_{t\to\infty} t \cdot \cos\theta_s}{\operatorname{sign}\theta_s \cdot \lim_{t\to\infty} t \cdot \sin\theta_s + d} \tag{7}$$

$$= \begin{cases} \pm \lim\limits_{t\to\infty} t & \text{if } \theta_s = 0 \\ -f_s / \tan\theta_s & \text{otherwise} \end{cases} \tag{8}$$

$$y_s' = 0 \tag{9}$$

Side:

$$x_s' = f_s \frac{-\tan\theta_e \cdot \cos\theta_s}{\tan\theta_e \cdot \sin\theta_s + 1}, \quad y_s' = 0. \tag{10}$$

The image size at this point depends on the focal length f_s. In a real camera, the image size is the sensor size.

3.4 Display

The actual display occurs on a physical display device, such as a computer monitor, tablet or phone. We model this by magnifying the image from sensor size to display size by the factor M. The viewer sits straight in front of the display at a viewing distance v; see Fig. 3 (right). The transformation from projected simulation coordinates to viewing coordinates is simply:

$$x_v = M \cdot x_s', \quad y_v = M \cdot y_s', \quad z_v = v. \tag{11}$$

3.5 Predicted Perceived Angle

We now have all the elements needed to compute the angle the viewer will perceive according to the extended retinal hypothesis. In particular, as explained by the vision literature (e.g. Sedgwick [1991]), the perceived total angle is the angle between the lines from the viewer position (at the origin of the viewer coordinate system) to the front and side vanishing points; see Fig. 3 (right). The angle formed by connecting a viewpoint to these two vanishing points is assumed to be the same for any viewpoint, because this is true by definition for veridical vanishing points. Specifically:

$$\alpha_{\text{front}} = \arctan |x_{v,\text{front}}| / z_{v,\text{front}} \tag{12}$$

$$= \begin{cases} 90° & \text{if } \theta_s = 0 \\ \arctan \dfrac{M \cdot f_s}{v \cdot \tan|\theta_s|} & \text{otherwise} \end{cases} \tag{13}$$

$$\alpha_{\text{side}} = \arctan |x_{v,\text{side}}| / z_{v,\text{side}} \tag{14}$$

$$= \arctan \left(\frac{M \cdot f_s}{v} \cdot \left| \frac{c \cdot \tan\theta_e \cdot \cos\theta_s}{c \cdot \tan\theta_e \cdot \sin\theta_s + d} \right| \right). \tag{15}$$

If θ_e and θ_s have opposite signs, the perceived angle is:

$$\alpha_{\text{total}} = \alpha_{\text{front}} + \alpha_{\text{side}}. \tag{16}$$

If θ_e and θ_s have the same sign, the perceived angle is $\alpha_{\text{total}} = 180° - \alpha_{\text{front}} + \alpha_{\text{side}}$.

Both faces of a corner must be visible for observers to make informed angle judgments. A discontinuity arises at eccentricity angle $\theta_e = 0$ because the side face of the corner flips between facing left and facing right.

4 Experimental Design

We conducted two psychophysical experiments to determine how the image distortions in typical street-level IBR applications are perceived:

1. An angle-matching experiment that tested the predictions of the scene and extended retinal hypothesis. The results allow us to predict the magnitude of perceived distortions in novel scenes.
2. A rating experiment that determined the subjective acceptability of angle distortions. The results allow us to determine which perceived angle distortions are acceptable and which are not; we will use this subsequently for our guidelines (Sec. 8).

For certain applications (e.g., navigation or sightseeing), one may be interested in the overall perception of shape. Since shapes involve many angles and different contextual priors, we believe it is likely to lead to more acceptable perceived distortions. We opted to study angles, which is a stricter criterion with a clear quantitative task, used before by Watt et al. [2005]. An analogous quantitative experiment for shape perception would not be possible.

4.1 Stimuli

We simulated the typical unstructured lumigraph workflow in a fully synthetic, offline rendering pipeline to maximize control and

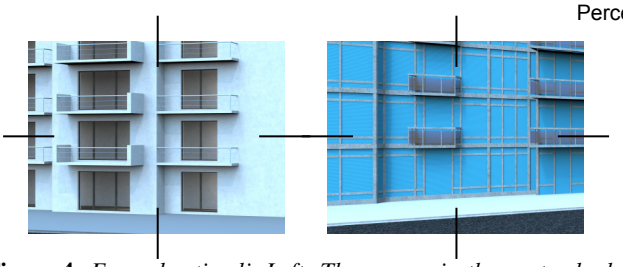

Figure 4: *Example stimuli. Left: The corner in the center looks undistorted. Right: the corner looks like an obtuse angle. The corners participants are asked to look at are indicated by the crosshairs (or by a blinking red dot in the experiment).*

image quality. Synthetic stimuli allow us to create depth variations of the same façade, and to specify exact and consistent values for eccentricities and capture camera positions across different façades, as well as to create images without clutter (e.g., cars, lamp posts, trees), and avoiding the presence of other cues in the stimuli.

Three façades (Figs. 4 and 5) were created with 90° convex corners at eccentricity angles of $\theta_e = \pm 7.1°$ and $\pm 32.0°$. Three balcony depths were created for each façade: the distances from the front to the back of the balconies were 0.33, 0.67, and 1 m. The façades were lit by a sun and sky illumination model to create realistic images.

The stimulus images were created using the PBRT offline renderer [Pharr and Humphreys 2010] with high-quality parameters. A wide-angle frontoparallel image of three realistic façades was rendered from a distance of 40 m. The resulting images were then projected onto a single plane. Details are provided in the supplemental material. The same stimuli were used for both experiments. Fig. 5 provides an overview of the stimulus creation parameters.

4.2 Hypotheses

The extended retinal hypothesis (Eq. 16) can be used to predict the perceived angle of corners in our stimuli. Fig. 7(a) shows the extended retinal hypothesis prediction for the viewer at the center of projection. The angle perceived when viewing a 90° corner is plotted as a function of simulation angle. Different colored lines represent the predictions for different eccentricity angles. Each corner has a constant eccentricity angle in a given captured image: e.g., the corner on the left in Fig. 5 has a negative eccentricity angle $\theta_e = -32°$ relative to the blue capture camera at the bottom, so it is represented by the dark blue line in Fig. 7(a). As the simulation camera shown in green rotates around that corner from left to right, the simulation angle θ_s increases and the predicted perceived angle gradually decreases towards the correct value. When the eccentricity angle is positive, the predicted perceived corner angle increases with increasing simulation angle. Thus, the extended retinal hypothesis predicts a two-way interaction between simulation angle and eccentricity angle.

4.3 Experimental Setup

We used four types of displays that differed substantially in size: screen diagonals ranged from 3.5" to 55". Cooper et al. [2012] found that users have fairly consistent preferred distances for viewing photographic prints of different sizes, given by

$$\textit{viewing distance} = 1.3 \cdot \textit{image diagonal} + 253\,\text{mm}. \qquad (17)$$

We assume the same formula applies for viewing display devices, so we used it to determine the viewer's distance from each device. The devices were held upright (PC, TV) or at an angle of about 45° (phone, tablet) by fixed stands, and the viewer was positioned at the specified distance. More details of the experiment implementation are in the supplemental material and the video.

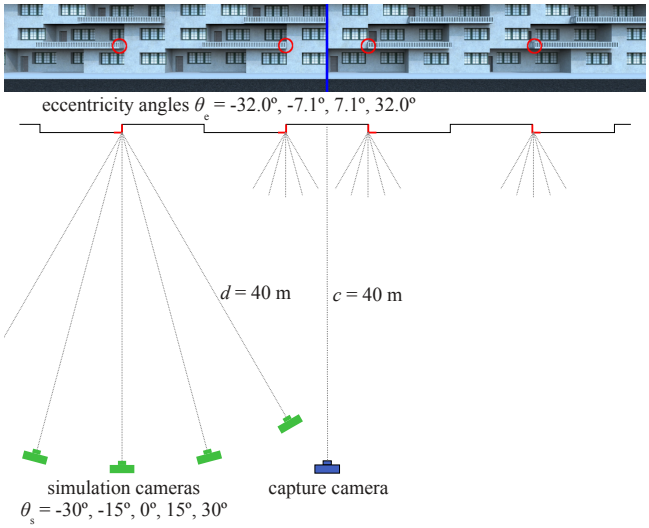

Figure 5: *Wide-angle image (cropped) from the capture camera (optical axis in blue, corners in red), and stimulus creation parameters.*

4.4 Participants

It is common practice in vision science to use a relatively small number of participants who are tested quite extensively (e.g., Ernst and Banks [2002]). We followed this practice by testing six participants extensively (on average 7.5 h for a total of 3072 measurements each). In so doing, we were able to test many experimental conditions and thereby put the hypotheses to rigorous test. The six participants were 20–32 years old; three were female. They all had normal or corrected-to-normal vision, and were paid for their participation. They all participated in both experiments and were naïve about the experimental hypotheses. The tasks in these experiments are quite involved, so we needed to ensure that participants understood the instructions, and that they could give reasonably consistent responses. To do this, candidates passed a pre-test, and only those satisfying comprehension and consistency requirements were invited to perform the entire test. Details of this pre-test are given in the supplementary material.

4.5 Experimental Procedure

In both experiments, all stimuli – five simulation angles, four eccentricity angles and three façades, each with three balcony depths – were presented twice. In addition, three particular stimuli were presented eight times to allow us to assess the consistency of responses. Thus, there were 384 stimulus presentations for each of the four display device types in each experiment. The order of stimuli presented was randomized. In addition, we used within-subject counterbalancing. There were two experiments (hinge setting and rating) and four devices, giving eight conditions. We further split these each into three parts, giving 24 parts. These 24 parts were randomized and split into separate sessions – usually four 2-hour sessions. Participants took breaks whenever they needed. At the start of each part, participants were given extensive instructions. They were shown examples of stimuli without angle distortions and with extreme angle distortions (Fig. 4).

5 Experiment 1: Hinge Angle Matching

The first experiment determined how a 90° corner is perceived after it undergoes distortion due to the position of the simulation camera. Participants indicated the perceived angle of a corner by setting a real convex hinge to look the same (see Figure 6). We used a real

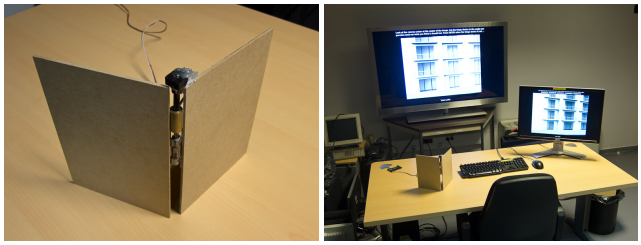

Figure 6: *The hinge device (left) and our experimental setup (right).*

Source	df	SS	MS	F	p	Sig	η^2_G
Display Device (Dev)	3	3572	1191	7.939	0.0021	⋆	0.105
Simulation Angle (Sim)	4	777.6	194.40	7.102	0.000995	⋆	0.024
Eccentricity Angle (Ecc)	3	56020	18673	50.05	4.78e-08	⋆	0.648
Façade Depth (Dep)	2	18788	9394	61.84	2.34e-06	⋆	0.382
Dev×**Sim**	12	56.3	4.692	0.551	0.872		0.001
Dev×**Ecc**	9	462.5	51.39	1.362	0.234		0.014
Dev×**Dep**	6	209.0	34.83	1.647	0.169		0.006
Sim×**Ecc**	12	17339	1444.9	11.39	1.92e-11	⋆	0.363
Sim×**Dep**	8	110.9	13.87	1.228	0.308		0.003
Ecc×**Dep**	6	2594	432.4	9.135	1.03e-05	⋆	0.078
Dev×**Sim**×**Ecc**	36	683.2	18.98	2.142	0.000592	⋆	0.022
Dev×**Sim**×**Dep**	24	174.7	7.281	0.884	0.622		0.005
Dev×**Ecc**×**Dep**	18	428.3	23.80	1.933	0.0225	⋆	0.014
Sim×**Ecc**×**Dep**	24	436	18.17	1.745	0.0268	⋆	0.014
Dev×**Sim**×**Ecc**×**Dep**	72	785	10.899	1.224	0.121		0.025

Table 1: *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 $\alpha = 0.05$, where $*$ means the factor is significant and generalized η^2 effect size.*

hinge, similar to Watt et al. [2005], because images of a hinge may themselves be affected by perceptual distortions.

Participants were shown images from our pool of stimuli (Sec. 4.1) in random order. The intended corner, which was always in the center of the image, was briefly indicated by a blinking dot. Participants then rotated the hinge device about the vertical axis until the slant of one side of the hinge appeared to match the side of the corner in the image. They then opened or closed the hinge until the perceived angles of the hinge and corner were equal. They pressed a keyboard button to indicate that they were satisfied with the setting, the setting was recorded, and the experiment proceeded to the next trial. This procedure is illustrated in the accompanying video.

5.1 Results and Discussion

The angle-matching experiment took 4.3 hours on average to complete, or about 10 seconds per trial. We used the eight repeated presentations of three stimuli to assess response consistency. Those data showed that participants were self-consistent (average standard deviation within-subjects was 5.3°). The data from different participants were also quite similar (average standard deviation between-subjects was 8.7°), so we averaged across participants to create the data figures shown here. Similarly, we did not observe systematic differences across display devices (average standard deviation between devices was 6.3°), so we averaged over devices. See supplementary material for the data from individual participants and devices.

Figure 7 shows the predicted (a), and observed (b–e) hinge angle settings. Each panel plots angle setting as a function of simulation angle; different colored lines represent different eccentricity angles. The colored lines in panel (a) represent the predictions for the extended retinal hypothesis and the horizontal dotted lines at 90° the predictions for the scene hypothesis. The results are qualitatively quite similar to the predictions of the extended retinal hypothesis. For positive eccentricity angles, the extended retinal hypothesis predicts that increases in simulation angle will yield increases in perceived angle. For negative eccentricity, the prediction is the opposite. All of the data are consistent with those predictions. However, the range of perceived angles is more compressed than predicted by the retinal hypothesis, which suggests an influence of the scene hypothesis. This partial influence of the scene hypothesis is greater for small façade depths (Figure 7, b–d).

To determine which effects are statistically significant, we performed a repeated-measures analysis of variance (ANOVA) on the data with simulation angle, eccentricity angle, display device and façade depth as factors. The results are shown in Table 1. There were statistically significant main effects of all factors, but only eccentricity angle and façade depth have a large enough effect size ($\eta^2_G > 0.26$) to result in noticeable differences in perceived angle [Bakeman 2005]. There was also a significant and large two-way interaction between simulation angle and eccentricity angle, as predicted by the retinal hypothesis.

The main effect of façade depth reflects the fact that shallower façades tend to be perceived as having smaller angle deviations than deeper façades as can clearly be seen in Fig. 7 (b–d). This is because deeper façades reduce the uncertainty in the retinal angle cues, allowing the retinal hypothesis to dominate more strongly over the scene hypothesis. Previous work has demonstrated that the perceived distortions do not become more objectionable with even greater depths [Vangorp et al. 2011].

The retinal hypothesis predicts that perceived angles deviate more from 90° as the simulation camera moves laterally from the capture position (see Fig. 2, top center); i.e., it predicts greater perceived angles when the simulation and eccentricity angles are large and have the same sign.

In a follow-up experiment, we presented stimuli created from real photographs in the same experimental setting. Five out of the 6 original participants were available for this follow-up experiment. We used the PC display only. The follow-up data with real images (available in the supplemental material) were very similar to the data with synthetic images. This validates our use of synthetic stimuli in the main experiment.

The retinal hypothesis (Eqs. 12–16) predicts different results for the different devices because the viewing distances relative to the COP differed across devices. But we did not observe systematic differences across devices (see data in supplemental material). Our explanation for the lack of a device effect is that the effect of distance from the COP is overshadowed by the compression towards 90° due to the familiarity of cube-like shapes [Yang and Kubovy 1999, Perkins 1972]. For this reason, Fig. 7 (a) plots the predictions of the retinal hypothesis with the viewer at the COP.

6 Experiment 2: Angle Rating

The second experiment was designed to determine the acceptability of various angle distortions. We asked participants to indicate how acceptable a given corner was as a simulation of a 90° corner. Participants were shown the same images as in Experiment 1, again in random order. They rated how close the indicated corner in each image looked to a right angle on a 5-point scale where 1 to 5 corresponded to “perfect”, “close enough”, “kind of”, “not really”, and “no way!”. Participants entered each rating using a numerical keypad and confirmed the entry by pressing “Enter” (see Fig. 8). The next stimulus then appeared and the process repeated.

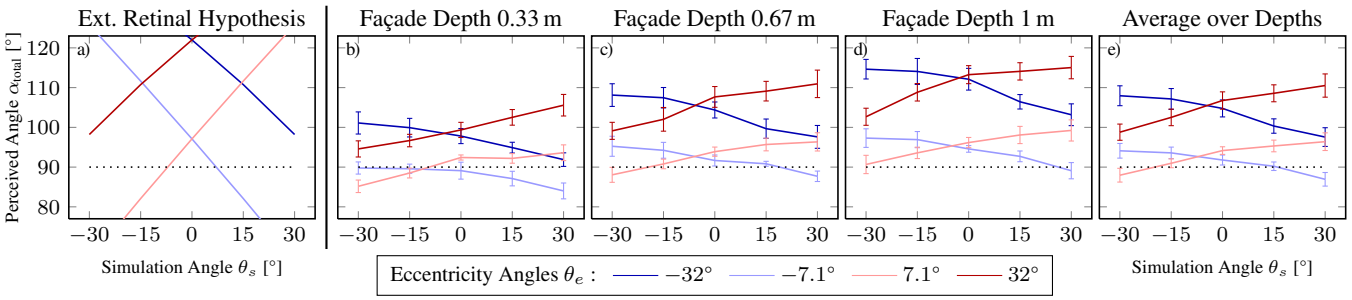

Figure 7: *Perceived angle predictions by the extended retinal hypothesis (a), compared to the angle-matching results for different façade depths (b–d) and averaged over all depths (e). Each line corresponds to an eccentricity angle; error bars indicate one standard error of per-subject means above and below the mean as a measure of between-subject agreement. The dotted line at 90° represents the scene hypothesis.*

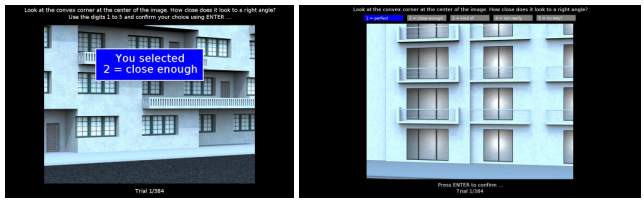

Figure 8: *Rating screens on the phone (left) and TV (right).*

Source	df	χ^2	p	Sig	W
Display Device	3	8.40	0.0384		0.467
Simulation Angle	4	9.73	0.0452		0.406
Eccentricity Angle	3	10.2	0.0169		0.567
Façade Depth	2	10.3	0.0057	⋆	0.861

Table 2: *Results of our four separate repeated-measures Friedman tests on the angle-rating data. The columns list: sources of variance, their degrees of freedom (df), χ^2-statistic, p-value, significance code for the Bonferroni-corrected significance level $\alpha = 0.0125$, and Kendall's coefficient of concordance W indicating effect size.*

In a brief training session, we encouraged participants to use the full range of the scale and to distribute their responses uniformly. In the instructions, we showed examples with "almost no" and "severe" angle distortions to help them understand the scale, but to avoid bias we did not tell them which ratings to assign for those stimuli.

6.1 Results

Medians of ratings on a 5-point scale are not very informative, so we use interpolated medians [Revelle 2008]. This summary statistic only assumes equal distances between the levels and uniformly distributed observations within the levels [Zar 2010], which are weaker assumptions than the ones required for computing means.

The rating experiment took on average 3.1 hours to complete on all four display conditions, or 8 seconds per trial. The within-subject reliability of ratings is again fairly good over time, with a quartile variation coefficient of less than 0.65 for all participants and devices. As is common with rating scales, some participants gave higher or lower ratings overall than others. This does not pose a problem, however, because all participants encountered all conditions, so we could assess the effects of interest uncontaminated by differences in overall ratings.

To determine which effects are statistically significant, we performed repeated-measures Friedman tests on the ratings separately for the factors simulation angle, eccentricity angle, display device, and façade depth. The results are shown in Table 2. There was only a statistically significant main effect of façade depth with a large effect size ($W > 0.5$).

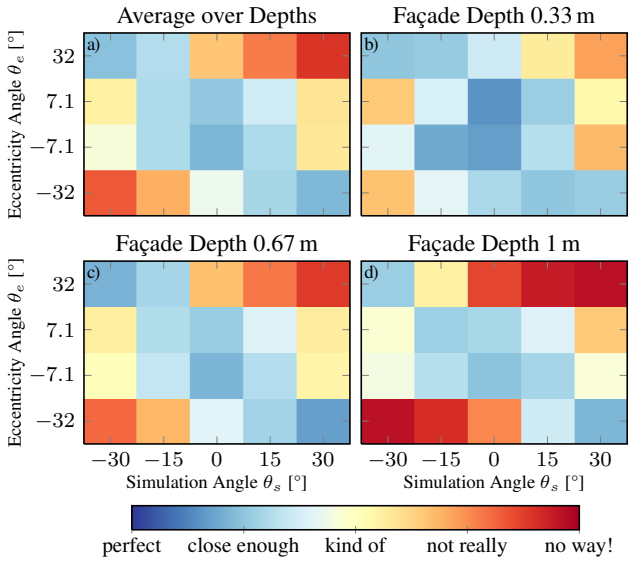

Figure 9: *Rating results showing interpolated medians across all participants: (a) across all façade depths, and (b–d) ratings for different façade depths.*

Figure 9 shows the results across participants. In each panel the abscissa and ordinate are respectively simulation angle and eccentricity angle, and colors represent different interpolated medians of ratings. The upper left panel shows the data across façade depths. The most unacceptable stimuli were in the lower left and upper right corners of these plots. Those stimuli have large simulation and eccentricity angles of the same sign. The most acceptable stimuli were in the middle of the plot – where the simulation and eccentricity angles are small in magnitude – and the upper left and lower right where the large simulation and eccentricity angles are opposite in sign. A comparison with Figure 7 shows that the most unacceptable stimuli were generally those that had perceived angles most different from 90°. The other three panels show the data for different façade depths. The largest and most systematic effects were observed for the largest depth. In supplemental material are the plots for different devices, which show that the ratings are quite consistent across display devices. We also performed a follow-up experiment with real stimuli, which were consistent with the synthetic data; results are in the supplemental material.

7 A Predictive Model for Perspective Distortion in Street-Level IBR

Our primary goal is to develop a *predictive model* of perspective distortion based on existing hypotheses and our experimental results.

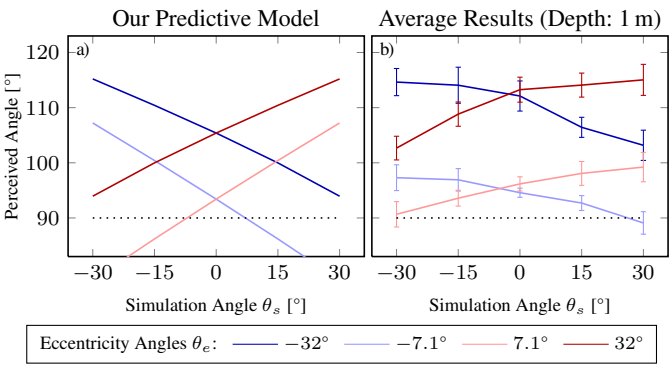

Figure 10: *Our predictive model (a) compared to the hinge angle-matching results (b), both for façade depth of 1 m. Each line corresponds to an eccentricity angle; error bars indicate one standard error about the mean. The dotted line at 90° is the scene hypothesis.*

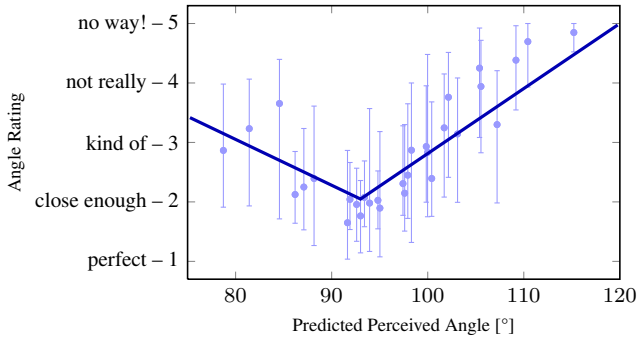

Figure 11: *The best piecewise-linear fit from perceived angles, as predicted by our model, to the ratings in our experiment. Each circle indicates an interpolated median rating for all stimuli with a given predicted angle. Error bars are quartiles. The best piecewise-linear fit to all data points (Eq. 20) is overlaid as solid lines.*

The results of the angle-matching experiment fell in-between the predictions of the retinal and scene hypotheses: i.e., as predicted by the retinal hypothesis, corners were often perceived as specifying angles different than 90°, but those perceived angles were closer to 90° than predicted by that hypothesis. We did not observe the significant effect of display device that was predicted by the extended retinal hypothesis. Previous work suggests that observers perceive less image distortion in images of familiar or cube-like shapes [Yang and Kubovy 1999, Perkins 1972]. However, the depth of the façade affected the results, with greater depths yielding results closer to the retinal hypothesis. A useful predictive model must incorporate these effects. To create such a model, we found best-fitting weights for the retinal and scene predictions for each depth value. The linear contribution of depth value is given by:

$$\text{fall-off}(\textit{depth}) = \begin{cases} y_0 & \text{if } \textit{depth} < \textit{depth}_0 \\ y_1 & \text{if } \textit{depth} > \textit{depth}_1 \\ y_0 + (y_1 - y_0) \cdot \frac{\textit{depth} - \textit{depth}_0}{\textit{depth}_1 - \textit{depth}_0} & \text{otherwise.} \end{cases} \tag{18}$$

where $0 < \textit{depth}_0 < \textit{depth}_1$ and $0 < y_0 < y_1 < 1$ are fitted to the data such that the prediction is not forced to follow the scene hypothesis completely at $\textit{depth} = 0$ or the retinal hypothesis at large *depth*.

The corner angle α predicted by the model then becomes:

$$\alpha = \alpha_{\text{total}} \cdot \text{fall-off}(\textit{depth}) + 90° \cdot (1 - \text{fall-off}(\textit{depth})), \tag{19}$$

where α_{total} is defined in Eq. 16. In Figure 10, (a) plots the predictions of our model and (b) shows the experimental results for comparison. The fit to the angle measurement data can be improved by taking into account a "flattening" effect due to other cues (see supplemental material), but we did not adopt it because it adversely affects rating predictions, which are more important for applications (Section 8).

8 Guidelines for Street-Level IBR

Our experiments provide a wealth of data that can be used to predict distortions and provide guidelines for street-level IBR. We now investigate the correlation between perceived angles and ratings, and show that it allows us to develop a predictor of distortion acceptability. We then discuss how our experimental results provide two guidelines for street-level IBR: how to use the predictor for IBR and how to specify capture positions to minimize distortion. For both guidelines, the position and orientation of a novel simulation camera must first be converted to the parameters θ_s, θ_e and d used in our predictive model (see supplemental material).

8.1 Relating Angle Measurements and Ratings

We use the results of our experiments to determine the amount of perceptual distortion that should be considered acceptable in real-world applications. The data of the angle-matching and rating experiments are generally consistent with one another. This is evident from Figure 12: (a) shows the deviation of perceived angle from 90° and (b) shows the ratings for the same stimuli. The plots are qualitatively quite similar.

Figure 11 shows the consistency in another way by plotting observed ratings as a function of the predicted perceived angle. As expected, the lowest (most acceptable) ratings are close to 90° and the ratings increase (become less acceptable) for greater and smaller angles. We fit a piecewise-linear function to this plot (represented by lines in the figure) and use this function to predict ratings from predicted perceived angles:

$$\textit{rating}(\alpha) = \begin{cases} 2.05 + 0.12 \cdot (\alpha - 93.0°)/° & \text{if } \alpha \geq 93.0° \\ 2.05 - 0.08 \cdot (\alpha - 93.0°)/° & \text{if } \alpha \leq 93.0°, \end{cases} \tag{20}$$

clamped to the range $[1, 5]$ used in the experiment.

8.2 Predicting Ratings in Novel Conditions

Naturally, we wanted to predict ratings at eccentricity and simulation angles that were not sampled in our experiments. Given the symmetry in the results of both experiments, we combined the data from the appropriate conditions (simulation and eccentricity angle both inverted) to obtain symmetric predictions with even lower variance. The experimental rating data could be interpolated directly (Fig. 12, b) or the perceived angle data could be interpolated and then mapped to ratings (Fig. 12, a). However, both approaches are limited to the range of simulation and eccentricity angles sampled in our experiments, and larger angles can occur in street-level IBR. Therefore, we used Equation 20 to expand perceived angle predictions to $\pm 60°$ (Fig. 12, d). The discontinuity at $\theta_e = 0°$ is preserved in this procedure.

8.3 Guidelines for Capture and Display

For novel camera positions and orientations that might occur in street-level IBR, we first convert to θ_e and θ_s, and then look up the rating from Fig. 12 (d). We describe the look-up procedure in supplemental material. In Sec. 9 we show how predictions for novel positions and orientations are used in a street-level IBR application.

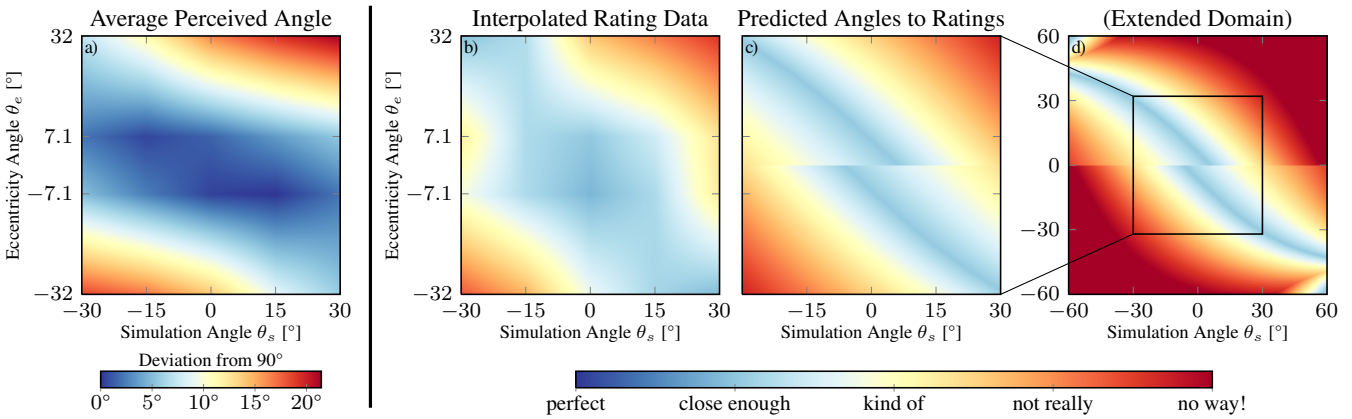

Figure 12: *(a) The deviation of averaged perceived angles (Experiment 1) from 90° has a similar pattern as (b), the interpolated medians of ratings from Experiment 2 (both linearly interpolated). (c) Perceived angles predicted by Eq. 19 mapped to ratings using Eq. 20, and (d) the same for an extended domain.*

We want to determine how positioning of the simulation camera affects the acceptability of the resulting images. The left side of Fig. 13 plots predicted acceptability for various camera positions and orientations relative to a façade with large depth range. In each panel, the capture camera (represented by a black icon) is facing the façade directly at a distance c of 10 m. The abscissa of each panel is the horizontal position of the simulation camera relative to the capture camera and the ordinate is the distance of the simulation camera to the façade. Different camera orientations are shown in different panels: θ_s is 0°, 15° and 30° from top to bottom. High and low acceptability is represented by blue and red, respectively. These plots were derived in part from Fig. 12: a column at the appropriate simulation angle in Fig. 12 (d) maps to a row in Fig. 13 for a simulation distance equal to the capture distance. These plots show how one can position the simulation camera to maintain acceptable imagery. The acceptable region does not vary much with simulation angle and it narrows as the simulation camera gets closer to the façade. The eccentricity and simulation angles are roughly opposite to one another in the acceptable zone.

The right side of Fig. 13 provides insight into how the density of capture cameras affects image acceptability for various simulation camera positions. The abscissas and ordinates of these panels are the same as those on the left side. The simulation angle θ_s is now always 0°. From the upper to lower panel, the capture cameras (black icons) have been positioned laterally at intervals of c, $c/2$, and $c/4$. As one can see, the density of capture cameras has predictable effects on the acceptability of the resulting images for a wide range of simulation camera positions. Our analysis, therefore, provides a principled method for determining how many captures are required for a given street scene and set of navigation paths. If the capture cameras are spaced out by distance c, getting from the region covered by one camera to the region covered by the next involves moving through regions of large distortion (yellow or red). An inter-camera distance of $c/4$ (third row) results in large regions of low predicted distortion, which is probably sufficient for street-level IBR.

9 Validation and Application

We developed a prototype interface for street-level IBR that uses the predictions described in the previous section. We first use this implementation to visualize the quality of a given path, and perform a study to validate our predictions. We then develop an application that guides users to stay in zones of acceptable quality.

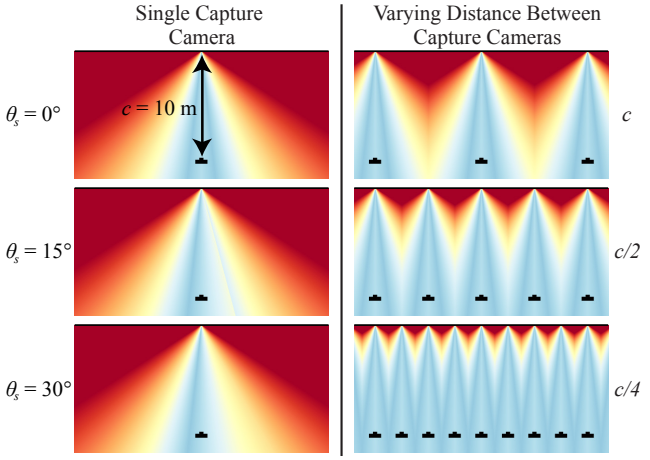

Figure 13: *Left: How simulation camera position and orientation affect predicted image acceptability. Each panel is a top view of acceptability as a function of the lateral position and distance of the simulation camera. From top to bottom, θ_s is 0°, 15° and 30°. Right: How positioning of capture cameras affects image acceptability. $\theta_s = 0°$. From top to bottom, capture cameras are positioned at intervals of c, $c/2$, and $c/4$.*

Visualization Our implementation reads a set of cameras registered using structure-from-motion [Snavely et al. 2006]. We assume that the cameras are directly facing the façade (i.e., a side camera in a commercial "capture car" for such applications). We fit a plane to the reconstruction which serves as the proxy for the façade.

We designed navigation paths in our interface, visualizing the predicted quality during the process. This is illustrated in Fig. 14. The interface shows the current simulated view and the inset shows a top view of the navigation path. The navigation path is color coded using the heat map of Fig. 12 (d) to show the predicted ratings. The accompanying video provides an animated example of navigation along the path.

While using our interface we noticed one significant temporal effect: motions along a path on which the perceived angle changes quickly are quite disconcerting. We can use our heat map of predicted angles to predict such paths in order to avoid them during navigation.

Figure 14: *Simulated views along a navigation path (see inset). The heat map of predicted ratings of Fig. 12 (d) is used as the color scheme to visualize predicted quality along the path. The view shown here is predicted to be bad, as indicated by the red line-of-sight in the inset (labeled 'S'). The capture camera is indicated by 'C'.*

Validation User Study To evaluate our predictions, we employed this interface in a user study. The goal of the study was to determine how well our predictions agree with user observations in a situation quite similar to street-level IBR: a navigation path with real stimuli. We used three datasets corresponding to the different scenes in Figures 1, 2 and 14. For each scene we provide three different paths: one path was predicted to be low quality (rating of 3–3.5), one medium quality (2.5–3), and one high quality (2–2.5). We created a web page with pre-recorded videos of each path. We instructed participants to look at a specific corner when it appeared in the middle of the screen (indicated by a red dot). Participants then rated the perceived distortion in the same manner as for Experiment 2: i.e., "how close to 90° does the angle look?". They chose a value on the same five-point scale (Sec. 6) for a total of nine paths (three paths for three scenes). The three videos of a scene were presented on a single page, and participants were asked to adjust their relative ratings between the three videos. This procedure is illustrated in the accompanying video. 91 people performed the study on their own computer screens. The results are summarized in Fig. 15, which plots observed ratings as a function of predicted ratings separately for the three scenes. The correlation between predicted and observed ratings was moderate ($r > 0.5$) for the first two scenes and strong ($r > 0.8$) for the third. Thus, the predictions were reasonably good despite the many differences between the experiments used to generate the predictions (static scenes with well controlled conditions) and this user study (unstructured dynamic scenes).

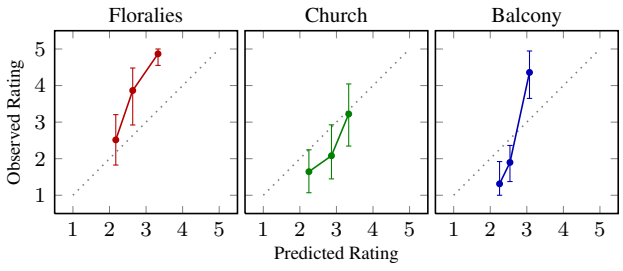

Figure 15: *Interpolated medians and quartiles for the three scenes used in the validation user study.*

Application We also developed an interactive application based on this implementation, illustrated in Fig. 16. The interface shows the simulated view and a top view of the scenario in the inset. The user starts viewing at a particular position, and then translates and/or

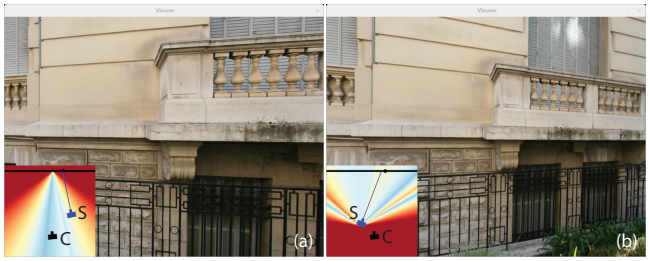

Figure 16: *Interactive navigation tool. (a) When the user translates, the inset shows predicted ratings for all camera positions keeping the orientation fixed. (b) When the user turns, the inset shows ratings for all possible camera orientations keeping the position fixed. The application restricts the user's motion to regions with acceptable predicted quality. Capture camera is indicated by 'C' and simulation camera is indicated by 'S' in the insets.*

rotates. If the user is translating (Fig. 16, a), the inset shows predicted ratings for all camera positions while keeping the same camera orientation. Fig. 13 (left) shows a similar visualization for three particular simulation angles. When the user turns (Fig. 16, b), the visualization shows ratings for all camera orientations keeping the camera position fixed. The user can translate and turn as they wish as long as they stay within the zone of "acceptable" quality. The application prevents the user from reaching a camera position or orientation that corresponds to a predicted rating higher than a threshold, and instead shows a blinking placeholder at the current camera position. This is illustrated in the accompanying video. We used a rating value of 3 as the threshold.

10 Discussion and Conclusion

In this work, we extended the retinal hypothesis from the vision science literature so that it could be applied to viewing façades in street-level IBR. We then performed an angle-matching experiment to measure the perceptual distortions that occur in such systems, which showed that perceived distortions are in between the competing retinal and scene hypotheses – depending on the façade depth. We also performed a rating experiment to determine the acceptability of the distortions measured in the first experiment.

We fit analytic functions to the perceived-angle data in order to create a predictive model of perceptual distortions in street-level IBR. By correlating the model to the rating data, we can predict the acceptability of different distortions. From this, we developed guidelines for acceptable navigation regions and capture positions for street-level IBR. Finally, we developed an application that predicts the quality associated with different navigation paths, and performed a confirmatory study that showed that our results generalize to realistic navigation systems.

In future work, we wish to extend our approach to more sophisticated IBR algorithms, e.g., to evaluate the usefulness of using more complex proxies for scene geometry. Our study is currently limited to the axis-aligned geometry of typical façades to keep the complexity under control. However, all geometry and viewing restrictions can be addressed by deriving the appropriate extended retinal hypothesis and predictive model from first principles as demonstrated in Sec. 3 and 7.

A very fruitful aspect of this project is the methodology we developed. We started with a common observation: our tolerance to changes in viewing angle when looking at texture-mapped façades. Vision science provided well-founded explanations of the processes underlying this phenomenon, but did not provide a directly applicable model. However, methodologies and geometric tools from

vision science allowed us to develop our extended hypothesis, and to design and run our experiments. In addition, we were able to develop useful applications of the theory and experimental results. We firmly believe that this geometric and experimental framework provides a solid basis for studying perception of IBR in more general settings, such as stereo viewing, and even transition artifacts that have recently drawn attention in the graphics community [Morvan and O'Sullivan 2009, Stich et al. 2011].

Acknowledgments

We thank the reviewers for their comments and all participants for their time. We also thank Jean-Pierre Merlet, Emmanuelle Chapoulie and Val Morash for their help, Charles Verron, Michael Wand, Karol Myszkowski and Holly Rushmeier for their comments, and Blend Swap users Sebastian Erler and *abab* for the façade models. This work was supported by the INRIA CRISP associate team, the EU IP project VERVE (www.verveconsortium.eu) as well as research donations by Adobe and Autodesk. Emily A. Cooper acknowledges support from the NDSEG Fellowship Program and National Science Foundation under Grant No. DGE 1106400.

References

ADAMS, K. R. 1972. Perspective and the viewpoint. *Leonardo 5*, 3, 209–217.

BAKEMAN, R. 2005. Recommended effect size statistics for repeated measures designs. *Behavior Research Methods 37*, 3, 379–384.

BANKS, M. S., HELD, R. T., AND GIRSHICK, A. R. 2009. Perception of 3-D layout in stereo displays. *Information Display 25*, 1, 12–16.

BUEHLER, C., BOSSE, M., MCMILLAN, L., GORTLER, S., AND COHEN, M. 2001. Unstructured lumigraph rendering. In *Proceedings of ACM SIGGRAPH 2001*, 425–432.

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.

DEBEVEC, P., YU, Y., AND BORSHUKOV, G. 1998. Efficient view-dependent image-based rendering with projective texture-mapping. In *Proceedings of EGWR '98*, 105–116.

ERNST, M. O., AND BANKS, M. S. 2002. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature 415*, 6870, 429–433.

GOLDSTEIN, E. B. 1987. Spatial layout, orientation relative to the observer, and perceived projection in pictures viewed at an angle. *Journal of Experimental Psychology: Human Perception and Performance 13*, 2, 256.

KOPF, J., CHEN, B., SZELISKI, R., AND COHEN, M. 2010. Street slide: browsing street level imagery. *ACM Transactions on Graphics 29*, 3, 96:1–8.

LUMSDEN, E. A. 1983. Perception of radial distance as a function of magnification and truncation of depicted spatial layout. *Attention, Perception, & Psychophysics 33*, 2, 177–182.

MORVAN, Y., AND O'SULLIVAN, C. 2009. Handling occluders in transitions from panoramic images: A perceptual study. *ACM Transactions on Applied Perception 6*, 4, 25:1–15.

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.

PIRENNE, M. H. 1970. *Optics, Painting and Photography*. Cambridge University Press.

REVELLE, W. 2008. *psych: Procedures for Psychological, Psychometric, and Personality Research*. R package version 1.0-42+.

ROSINSKI, R. R., MULHOLLAND, T., DEGELMAN, D., AND FARBER, J. 1980. Picture perception: An analysis of visual compensation. *Attention, Perception, & Psychophysics 28*, 6, 521–526.

SEDGWICK, H. A. 1991. The effects of viewpoint on the virtual space of pictures. In *Pictorial Communication in Virtual and Real Environments*, S. R. Ellis, Ed. Taylor & Francis, 460–479.

SHUM, H. Y., CHAN, S. C., AND KANG, S. B. 2006. *Image-based rendering*, vol. 2. Springer.

SMITH, P. C., AND SMITH, O. W. 1961. Ball throwing responses to photographically portrayed targets. *Journal of Experimental Psychology 62*, 3, 223.

SNAVELY, N., SEITZ, S. M., AND SZELISKI, R. 2006. Photo tourism: exploring photo collections in 3D. *ACM Transactions on Graphics 25*, 3, 835–846.

STEINICKE, F., BRUDER, G., AND KUHL, S. 2011. Realistic perspective projections for virtual objects and environments. *ACM Transactions on Graphics 30*, 5, 112:1–10.

STICH, T., LINZ, C., WALLRAVEN, C., CUNNINGHAM, D., AND MAGNOR, M. 2011. Perception-motivated interpolation of image sequences. *ACM Transactions on Applied Perception 8*, 2, 11:1–25.

TODOROVIĆ, D. 2008. Is pictorial perception robust? the effect of the observer vantage point on the perceived depth structure of linear-perspective images. *Perception 37*, 1, 106.

TODOROVIĆ, D. 2009. The effect of the observer vantage point on perceived distortions in linear perspective images. *Attention, Perception, & Psychophysics 71*, 1, 183–193.

VANGORP, P., CHAURASIA, G., LAFFONT, P.-Y., FLEMING, R. W., AND DRETTAKIS, G. 2011. Perception of visual artifacts in image-based rendering of façades. *Computer Graphics Forum 30*, 4 (Proceedings of EGSR 2011), 1241–1250.

VINCENT, L. 2007. Taking online maps down to street level. *Computer 40*, 118–120.

VISHWANATH, D., GIRSHICK, A. R., AND BANKS, M. S. 2005. Why pictures look right when viewed from the wrong place. *Nature Neuroscience 8*, 10, 1401–1410.

WALLACH, H., AND MARSHALL, F. 1986. Shape constancy in pictorial representation. *Attention, Perception, & Psychophysics 39*, 233–235.

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.

YU, J., MCMILLAN, L., AND STURM, P. 2010. Multiperspective modeling, rendering and imaging. *Computer Graphics Forum 29*, 1, 227–246.

ZAR, J. H. 2010. *Biostatistical Analysis*, 5th ed. Prentice Hall.