Quantifying reality of ultra-realistic 3-D displays — the effect of resolution and contrast

JOSEPH G. MARCH, University of Cambridge, UK
DOUNIA HAMMOU, University of Cambridge, UK
SIMON J. WATT, Bangor University, UK
RAFAŁ K. MANTIUK, University of Cambridge, UK

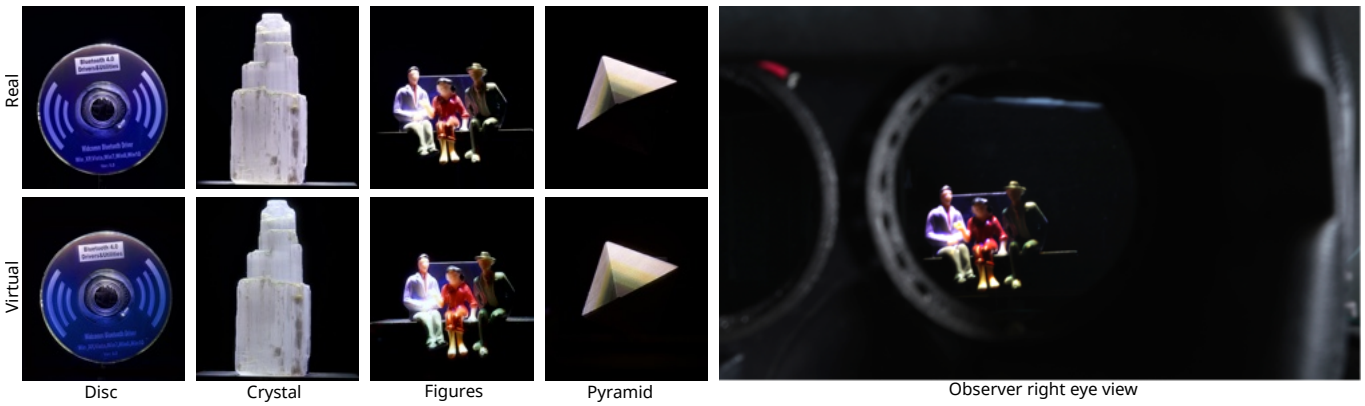


Fig. 1. Examples of the real and virtual versions of our stimuli scenes, photographed from the point of view of an observer. Virtual scenes were rendered at the highest resolution and contrast levels. Note that minor geometric differences between real and virtual scenes are caused by misalignment between the geometric display calibration and the nodal point of the camera; such inaccuracies were not visible to observers when viewing the display after calibration. The color pattern seen at the apex of the "Pyramid" was caused by an imperfect mixture of spectral components of our spectrally tunable light.

Developments in display technology mean that it may soon be possible to create stereoscopic 3-D imagery that cannot be distinguished from its real-world counterpart, i.e., it could pass the so-called Visual Turing Test. To measure how far current display technologies are from this ultimate goal, we built a custom ultra-realistic display — a high dynamic range (HDR) stereoscope, with accurate color and geometric calibration. Then, we used the display to measure how the reduction of display resolution and dynamic range (contrast) affects the sense of realism, relative to a real scene. These two display capabilities have improved greatly in recent years, and it is useful to understand what benefits further development might bring. The display resolution measurements showed that the realism of presented scenes starts dropping rapidly below ~60 pixels per degree, and the gains diminish above that resolution. Our contrast measurement showed a linear reduction in realism as the contrast was reduced. Furthermore, we found that the observers could discriminate between real and virtual scenes in less than half of the trials (when adjusted for the guess rate), demonstrating the ultra-realistic capabilities of our display. Our results provide a clear guide to potential gains in realism, and can inform display design and HDR tone mapping.

Authors' Contact Information: Joseph G. March, joemarch010@gmail.com, University of Cambridge, Cambridge, UK; Dounia Hammou, , University of Cambridge, Cambridge, UK; Simon J. Watt, , Bangor University, Bangor, UK; Rafał K. Mantiuk, rafal.mantiuk@cl.cam.ac.uk, University of Cambridge, Cambridge, UK.

SIGGRAPH Conference Papers '26, Los Angeles, CA, USA

CCS Concepts: • **Computing methodologies** → **Perception**; **Display Technology**.

Additional Key Words and Phrases: perception, perceptual realism, display technology, display resolution, image contrast, HDR, 3D displays, visual Turing test

ACM Reference Format:
Joseph G. March, Dounia Hammou, Simon J. Watt, and Rafał K. Mantiuk. 2026. Quantifying reality of ultra-realistic 3-D displays — the effect of resolution and contrast. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Papers (SIGGRAPH Conference Papers '26), July 19–23, 2026, Los Angeles, CA, USA.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3799902.3811097

1 Introduction

Display technology available today offers capabilities that match and sometimes exceed the limitations of the human eye. It is therefore becoming possible to create virtual images that are almost indistinguishable from real scenes [Borg et al. 2012; Masaoka et al. 2013; Zhong et al. 2021]. However, achieving such high fidelity is often impractical — it involves complex optical systems that cannot easily be miniaturized (e.g., for head-mounted displays), and requires excessive bandwidth and computation to deliver data for such displays. To build practical display systems, we therefore need to find the best balance of different display capabilities that manages system-specific requirements (form-factor, power consumption, etc.) alongside the least possible degradation in image appearance.

In this work, we examine how principal display capabilities, resolution and dynamic range (contrast), contribute to the realistic appearance of 3-D images. Our study employed four interrelated features aimed at better understanding how display capabilities might be appropriately balanced in practical contexts. First, we examined perceptual realism, rather than image quality. Realism is aligned with the ultimate goal of many 3-D display technologies, and allows the use of a real-world reference, which provides an objective and concrete benchmark against which to evaluate the effects of reduced fidelity. Second, we used scenes depicting real objects. Classical psychophysics stimuli (such as Gabor patches) are useful for determining the maximum possible sensitivity of the visual system to reductions in fidelity for a given display capability (e.g., resolution [Ashraf et al. 2025]). Reductions in fidelity may be less noticeable, and therefore less critical, in richer scenes more typical of 3-D graphics use cases, however. Third, we examined the effects of reducing the fidelity of individual display capabilities in 3-D images that were very-high-fidelity in most other respects. Compared to low-fidelity images, this is advantageous for determining the practical importance of display capabilities for realism in two ways. Low-fidelity images could mask the effects of improving a given display capability by placing a ceiling on how realistic the image can appear. For example, increasing contrast beyond a certain point may confer no further improvements if resolution or color gamut are insufficient for realistic appearance. Also, reducing the fidelity of a single display capability in low-fidelity images may have a larger effect on realism than would be the case when all other aspects are very high fidelity. Finally, we quantified the effects on perceptual realism of degrading resolution and contrast on a common, continuous scale. We obtained pairwise comparisons across different stimulus conditions, scaled under the Thurstone case V assumptions [Perez-Ortiz and Mantiuk 2017]. This provides a measure of the reduction in perceived realism on a continuous scale of just-noticeable-difference (JND) units. This allowed us to quantify how much perceptual realism is reduced by degrading fidelity not only at (or near) the maximum fidelity level needed to match human sensitivity, but also at lower-than-ideal levels (as used by many current devices). Moreover, this approach yields a common perceptual realism scale across all conditions, allowing the magnitude of effects of degrading unrelated display capabilities (resolution and contrast) to be directly compared.

Taken together, the features of our experiment allow us to understand how resolution and contrast contribute to a highly valid aspect of image appearance — perceived realism — under conditions that are relevant to practical use cases for 3-D graphics, providing useful information for display developers.

We measure the degradation of realism for two primary display capabilities: resolution and dynamic range. Resolution is perhaps the most salient display capability and has seen dramatic improvement over the last 20 years — from 0.35 M pixels for SD content (720×480) to 8.3 M pixels for UHD content (3840×2160). However, delivering high resolution is expensive, as it requires more power and more expensive components. We are therefore interested in understanding how this technological advancement translates into appreciable improvements in the perceived realism of displayed scenes. Display dynamic range (or contrast, which is equivalent) is another key display limitation. If the dynamic range of the content exceeds the display dynamic range, the former needs to be compressed via the process known as tone mapping. Optical see-through augmented reality (OST-AR) displays are particularly affected by limited dynamic range as their minimum luminance is elevated by the light from the environment. We seek to understand how contrast reduction translates into a degradation in perceived realism. Furthermore, we examined the effects of degrading both display capabilities at three luminance levels, 10× apart. Because the eye's sensitivity changes with luminance, it is possible that different display capabilities are needed depending on luminance. In order to examine the effects of degrading display capabilities in otherwise highly realistic stimuli, our experiments used a custom-built ultrarealistic high-dynamic-range haploscope (similar to that used by Zhong et al. [2021]). This display can reproduce nearly all static visual cues with very high fidelity, including color, spatial detail, and depth. The display also allows us to switch seamlessly between viewing real and virtual objects during experimental trials, allowing direct comparisons with a real-world reference (in fact, we demonstrate that the display can almost perfectly reproduce some of the 3-D objects, to the point that the virtual reproductions are confused with real objects). These properties allow us to examine the effects of degrading a single display capability on perceptual realism, and to anchor our perceived realism measurements against a perfect, ground-truth reference.

The experimental data and supplementary information can be found on the project webpage[1]. The light fields captured in the project are available in the associated dataset[2].

2 Related work

2.1 Display capabilities that meet human visual acuity

As introduced above, one approach to specifying display capabilities is to determine the fidelity required to meet or exceed the sensitivity of the human visual system on a given dimension.

Hopper [2000] attempt to characterize the display capabilities necessary to reach the visual acuity exhibited by observers with 20/20 vision. Similarly, Curry et al. [2003] present a more in-depth characterization of the display capabilities required to reach the limits of human visual acuity. Of particular interest is that both studies take into account the cost/benefit analysis of the display technologies required to reach such a threshold. This important consideration is often ignored in academic work, but is highly important to display developers. Recently, Ashraf et al. [2025] showed that the display resolution required to match the resolution of the human eye can be as high as 96 pixels-per-degree (ppd) for foveal achromatic vision. Considering display capabilities more generally, Hamilton et al. [2022] present a theoretical analysis of the spatial and directional resolution requirements necessary for a light-field display to be indistinguishable from viewing an equivalent real scene (i.e., to pass the so-called Visual Turing Test). This study gives a conservative (upper-bound) estimate of these requirements by disregarding the image statistics of natural scenes, instead focusing on the resolutions necessary to reach the peak contrast sensitivity

[1]Project web page: https://www.cl.cam.ac.uk/research/rainbow/projects/quantifying_reality/
[2]Dataset: https://doi.org/10.17863/CAM.129937

threshold of an observer. Wang et al. [2024] propose a visual clarity metric for near-eye displays, such as VR headsets. They propose to use visual acuity charts (e.g., a tumbling "E" chart) shown in a virtual environment to determine headset resolution. In contrast to the physical resolution limit, often given in pixels-per-degree, visual acuity accounts for other factors, such as the effective loss of resolution due to the headset's optics. Our measurements are performed on a high-quality haploscope that is free from aberrations and, therefore, can be considered device-independent.

2.2 Presenting highly realistic stimuli

Previous work has demonstrated that imagery presented on an electronic display can appear visually indistinguishable from that viewed in the real world but only under certain, constrained viewing conditions. Meyer et al. [1986] made the first attempt at reproducing reality. They constructed the famous Cornell box scene and used it to show that a computer-generated image can appear indistinguishable from a real scene, when viewed through a view camera. A commercially available CRT display was used to present the virtual content and the overall aim of the work was to produce computer-generated imagery with believable virtual lighting. The visual cues of the real and virtual scenes were degraded using a custom view camera with a Fresnel lens and ground glass. This places the study conceptually closer to photorealism than true perceptual realism. Borg et al. [2012] conducted a Visual Turing Test experiment in which they successfully replicated Meyer's results without requiring participants to view the stimuli through a viewfinder. Participants were shown either a real object or a display, both presented through a small aperture in a 2 m long box, with the stimuli viewed monocularly. Because the display could not achieve the necessary dynamic range, the authors instructed participants to observe the images from 10 cm away from the aperture in a moderately lit room (50 lux). Under these conditions, the glare in the eye and visual adaptation effectively masked the display's insufficiently dark black level. This limitation restricts the generalisability of the results to more common real world viewing scenarios. Zhong et al. [2021] used a high-dynamic-range multiple-focal-planes stereoscopic display, similar to ours, to present virtual scenes that reproduced a large number of real-world visual cues near-correctly. Similar to Borg et al. [2012], they directly examined whether participants could distinguish between real stimuli and their virtual counterparts. However, Zhong et al. operationalized the Visual Turing Test by using an 'oddity task', in which observers were shown three scenes and asked to select the one that was different (two real and one virtual, or one real and two virtual). This is a strong test of whether virtual and real stimuli are truly indistinguishable because it eliminates subjective criteria for what constitutes 'real' appearance. However, it is also a very difficult test to pass in practice because any small stimulus artifacts can create a discriminable difference between real and virtual stimuli, allowing the odd-one-out to be identified even if real and virtual stimuli appear equally realistic. Zhong et al. found that real and virtual stimuli were confused in 44% of cases (adjusted for chance). However, this result was obtained for a single scene that was shown at a very low luminance level (the peak luminance was 2 cd/m^2). In this work, we explore a much larger range of luminance, reaching over 1000 cd/m^2, for multiple scenes, and measure the effect of two key degradations of display capabilities.

2.3 Impact of display capabilities on perceived realism

Several previous studies have characterized how display capabilities affect the degree of perceptual realism, though not under the same conditions as the present study.

Masaoka et al. [2013] measured the relationship between spatial resolution and degradation of realism, similar to one portion of our experiment. The stimulus was either a real scene or an LCD display, both seen from 4.8 m through a synopter — a device that effectively removes binocular depth cues. In contrast, our stimuli were HDR 3-D scenes seen from 0.61 m, therefore delivering richer visual cues. We compare the results of their study with ours in Section 5.1. March et al. [2022, 2024] conducted studies assessing the importance of focus cue presentation on observer perceptions of realism when viewing content on a near-eye display. Their findings indicated that when viewing complex, small-scale scenes, presenting (near) correct focus cues provides a measurable but small enhancement to observers' perceptions of realism. Kim et al. [2024] used a computer-generated hologram (CGH) to assess the importance for perceptual realism of subtle ocular parallax cues, derived from eye movements. They found that ocular parallax meaningfully enhanced the realism of virtual scenes, albeit for large depth ranges (0 to 9.57 Diopters (D)), where the signal is particularly large. Their experiment also lacked important visual cues — the stimuli were viewed monocularly, lacking any binocular depth cues, and the resolution, dynamic range, and image quality were severely limited by the artifacts and limited capabilities of the holographic display. The study did not employ a real-world reference, which could help to interpret the magnitude of improvements in realism. Matsuda et al. [2022] studied the luminance level required for content to appear realistic and immersive in a custom-built high-dynamic-range VR headset. They concluded that most scenes require an average luminance level of over 100 cd/m^2 to maximise observer perceptions of realism, with some outdoor scenes requiring over 1000 cd/m^2. It should be noted that the scenes in this experiment included a variety of light sources that would be familiar to observers through real-world experience, including sunlight. It is unclear whether Matsuda et al.'s findings would generalize to scenes illuminated by arbitrary, non-natural illumination (such as those in [Zhong et al. 2021] and [Masaoka et al. 2013]), which would lack familiar luminance-level 'anchors'.

3 Apparatus

We used a custom-built high-dynamic-range multi-focal stereo display similar to the one used in [Zhong et al. 2021], but improved in several ways. The display, shown in Figure 2, is configured as a Wheatstone stereoscope, with separate left- and right-eye displays viewed via beamsplitters in front of each eye. Each eye's display has two focal planes (near and far). The images from near and far planes are combined using a 50T/50R beamsplitter (Figure 2(a)). Each focal plane is formed by a dual-modulation HDR display [Seetzen et al. 2004], in which a DLP projector (Acer P5535, color wheel removed) provides a backlight and an LCD panel (15.6" IPS 3840×2160 4K LCD LQ156D1JX02 with the original backlight removed) serves as a front

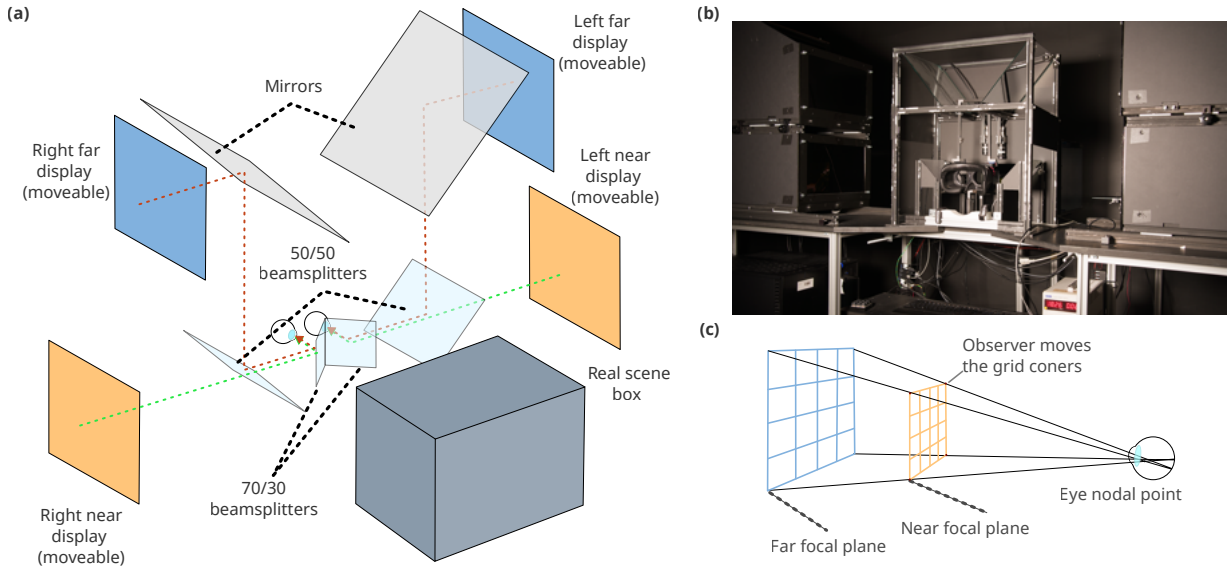


Fig. 2. (a) Schematic view of our HDR-MF-S display. (b) Photograph of the display. (c) Calibration process through which estimates of the positions of the nodal points of each observer's eyes were derived, to ensure accurate stereoscopic presentation.

color modulator. The projector-based HDR system lets us achieve high luminance levels up to 4 000 cd/m^{2}and high local contrast. The focal planes were positioned at viewing distances of 610 mm and 1092 mm from the eye, resulting in a resolution of the near plane of 120 ppd and Nyquist frequency of 60 cpd (cycles per degree). The far plane was used for geometric calibration (see Figure 2(c)), but was not used to show the stimuli.

The real scene was presented in a physical box, positioned beyond the Wheatstone beamsplitters (Figure 2a). The scene could be shown or hidden by switching on/off controllable spectrally-tunable lights (iQ-LED, Image Engineering) inside the real-scene box. The lights were set to simulate D65 standard illuminant. The real scene had to be small enough to fit within the field of view of our display and be illuminated in a controlled and reproducible manner. To satisfy these requirements, the box was 490×300×390mm, with the open front facing the observer as shown in Figure 3(a). The box was large enough to ensure that, when viewed through the display's aperture, only the back wall was visible to the observer. The inside of the real-scene box was coated in Vantablack© to reduce both the visibility of the back wall and secondary reflections during scene capture. Objects were placed on a holder constructed from matte black acrylic, which slotted into a groove in the floor of the box. These measures resulted in scenes in which one or more objects appeared to be floating against a black background, and any other part of the real-scene box was invisible to the observers.

Because the scenes had to be presented to observers in an arbitrary order, it was essential that the positioning of the real objects was reproducible, so that the (previously captured) virtual scenes and the real scenes precisely matched. The design of the real-scene box and mountings allowed objects to be swapped in and out as needed, while precisely preserving the position and orientation of the objects within the scene.

To achieve almost perfect geometric alignment, we represented our virtual scenes as 4-D light fields [Isaksen et al. 2000], which could be rendered from the viewpoint of each eye. To capture the images needed to produce these light fields, we used a bespoke camera rig depicted in Figure 3(a). Each image forming the light field was captured using a multi-exposure HDR merging technique [Hanji et al. 2020] to ensure sufficient dynamic range and negligible noise levels. Accurate color calibration was achieved by measuring a color checker (X-Rite Passport) with a spectroradiometer (JETI Specbos 1211 + Focus 3 mm imaging optics), photographing the same color checker with the camera, and then finding the root-polynomial color correction matrix [Finlayson et al. 2015]. To account for the spectral transmittance of the beamsplitters, we measured a reference white patch (2" Spectralon© White Balance and Diffuse Reflectance Target) with and without the beamsplitters and used the ratio of both spectra to correct the measured spectra of the color checker.

A pair of gantries allowed us to move the camera to any position on a rectangular plane perpendicular to the front of the real-scene box, as shown in Figure 3(a). The camera rig was positioned at 55 cm from the scene, in order to ensure that high spatial frequencies were captured.

We used a Sony A7R III camera with a ZEISS FE 1.8/55 (model SEL55F18Z) lens to capture images of the scene. We refer to this as the "scene camera". The pose of the scene camera was estimated using an image from an additional camera (IDS U3-3800CP Rev.2.2) mounted on top of the scene camera, which we refer to as the "pose camera". The pose camera captured images containing a checkerboard at a known location, which allowed us to estimate its pose. As the pose camera was mounted in a fixed position and orientation relative to the scene camera, the pose of the main scene camera could readily be estimated for each captured image. Real scenes

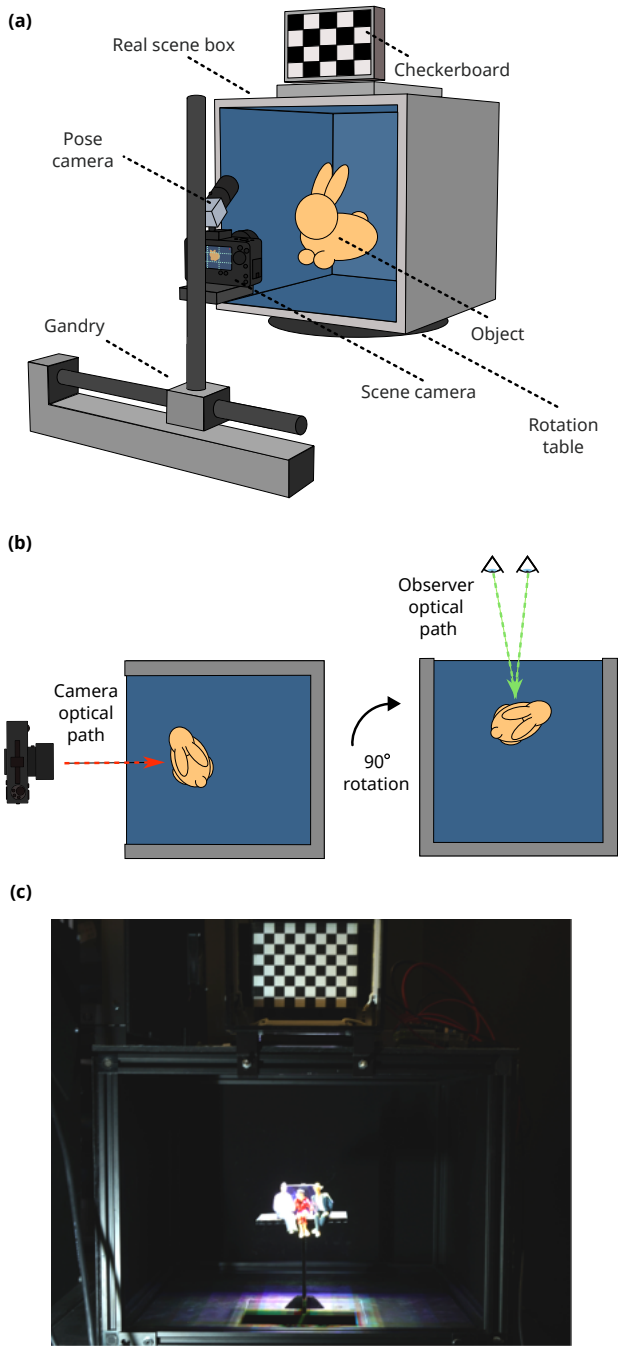


Fig. 3. (a) Schematic view of the light field capture rig used to generate the images used in our light fields. The scene camera moves to each position in a grid, and takes a multi-exposure photograph of the scene. (b) Overview of the rotation mechanism of the real scene box. (Left) The real scene box rotated to lie on the optical path of the camera for light-field capture, and (right) rotated to lie on the optical path of the observer for the experiment. (c) Photograph of a scene (Figures) positioned inside the real-scene box.

were captured by rotating the real-scene box 90 degrees, and photographing it from the side, as shown in Figure 3(b). This avoided introducing obstructions and optical distortions that placing the camera in the optical path of the display would introduce.

All camera systems, suffer from optical aberrations, which distort the captured image. We used a measured modulation transfer function (MTF) to compensate for the lens distortions. We estimated the MTF of our camera using the slanted-edge technique [Burns and Williams 2018] and a Modulated Sinusoidal Siemens Star chart (Image Engineering) as the target. The compensation was achieved by the Wiener deconvolution with the estimated MTF. To obtain a smooth MTF estimate, we fit a Gaussian function to our measurements. Because deconvolution using low MTF values results in the amplification of noise, we restricted the MTF values to be not smaller than 0.5.

4 Experiment

Our experiment aims to study observer perceptions of realism when viewing real and virtual stimuli, as a function of degradations in spatial resolution and dynamic range.

4.1 Stimuli

To allow our results to generalize well, our stimuli consisted of scenes containing objects with a wide variety of geometric and material properties, and of moderate size and depth. Figure 1 shows photographs of the four scenes, captured from the viewpoint of the observer's right eye. We selected scenes of simple ("Disc", "Pyramid") and complex geometry, with simple ("Figures") or complex textures, Lambertian ("Pyramid") and non-Lambertian materials. The objects in each scene were placed at a distance approximately equal to that of the near focal plane (610 mm). To avoid strong focus cues in our real scenes, we selected objects with a limited depth range (max 0.22 D, with the objects situated at a depth of 570 mm to 650 mm from the viewpoint of the observer). Each scene was presented at one of three illumination levels. The illumination levels were selected so that the median luminance of the object (excluding the dark background) was 0.7, 7, or 70 cd/m^2. Note that at the lowest illumination level, the dark portion of the object was below 0.05 cd/m^2 and at the highest illumination level, some pixels reached over 1000 cd/m^2. The luminance of the reference white patch (Spectralon, reflectance 98.8% at 550 nm) at 45° angle to the light and the eye was 5.24, 52.4, and 524 cd/m^2, respectively, for each illumination level.

To assess the importance of individual display attributes, we introduced two types of distortions when rendering our virtual stimuli: reduction of spatial resolution and of dynamic range.

Reduced spatial resolution: In order to simulate the impact of presenting the content on a display of lower resolution, we downsampled the images using a Lancosz ($a = 3$) filter, then upsampled them using a nearest-neighbour (box) filter. We used only integer resampling factors (2, 4, 8). This procedure most closely simulated images shown on displays of lower resolution, including aliasing due to the pixel structure. The effective low-pass filter for halving the resolution is shown in Figure 4. We report the resolution in units of pixels-per-visual-degree (ppd). Since ppd represents the sampling

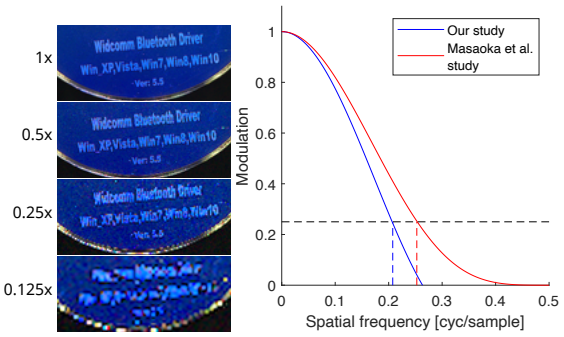


Fig. 4. Left: Example of the conditions with reduced resolution for the "Disc" scene. Right: The filter used in our study and that of Masaoka et al. [Masaoka et al. 2013] to halve the resolution. Dashed lines show the frequency at which the modulation was 0.25.

rate, the corresponding Nyquist spatial frequency represented in cycles-per-degree is half of the reported ppd value. This frequency corresponds to the point where the modulation of our resampling filters is 0 (see Figure 4).

Reduced dynamic range: While the dynamic range of content could be reduced using a wide range of techniques [Eilertsen et al. 2017; Reinhard et al. 2010], we wanted to focus on the most fundamental operation — contrast compression. Contrast compression can be achieved using a power function (gamma) on linear color values (BT.709 RGB color space), relative to a fixed luminance value [Mantiuk et al. 2015, sec. 5.2]:

$$Y_c\left(x,y\right) = w\left(\frac{X_c\left(x,y\right)}{w}\right)^{\gamma} \tag{1}$$

where $Y_c\left(x,y\right)$ is the output linear color value at coordinates (x,y) and for color channel c (RGB), X is the input linear color value, w is the measured median luminance of the light-field images, excluding the pixels that belong to the dark background. γ is a free parameter used to control the contrast of Y. The visual impact of this process on the "Pyramid" scene and its effect on an image histogram is shown in supplementary Figure S.4. Contrast compression will bring the intensity of color values closer to the selected luminance value w. This allows showing content on a display of reduced dynamic range. We did not clip color values to simulate the restricted dynamic range, as we did not want to introduce other confounding factors. We used a very large range of contrast manipulations in our experiment, with γ varying from 0.4 to 1, because we wanted to include the range that is relevant to optical see-through AR displays, whose dynamic range is significantly restricted when used in bright conditions.

4.2 Observers

23 observers (12 females, 11 males; mean age 25.6) took part in our experiment. All observers were naive to the purpose of our study. Prior to running the experiment, we tested observers' color and stereo vision using the Ishihara and Titmus Fly tests, respectively. All observers had normal stereo vision and normal color vision. Additionally, we tested observer visual acuity using the Snellen test. The Snellen test was performed at a distance of 3 m, such that each letter subtends 5 arcminutes at the observer's eye. The mean visual acuity over all observers was 0/16.3 (logMAR=-0.08). No participants were excluded. A more detailed analysis of the visual acuity of the observers is presented in the supplementary material. Observers were rewarded for their participation. In instances where observers were under the age of 18, their parents or guardians gave permission for their participation. The experimental procedure was reviewed and approved by the departmental ethics committee.

In order to account for differences in the separation of the eyes (interocular distance; IOD), and head position, across observers, and therefore to correctly present stereoscopic content, each observer was asked to complete a calibration procedure, which estimated the position of the nodal points of each eye. To do this, we showed a grid on each of the focal planes of our display (see Figure 2(c)). Observers were instructed to drag the corners of the grid presented on the near focal plane with a mouse to align them with those of the grid presented on the far focal plane. By tracing rays through the four corners of the grid, we could determine the position of the eye nodal point, as shown in Figure 2(c). This process was performed once per eye. We validated the procedure by comparing the IODs measured using a standard procedure (with a ruler) and using our method. To further validate the calibration, we showed observers two overlapping grids based on their estimated eye position. The algorithm used to determine eye nodal point locations is identical to that used in [Wernikowski et al. 2024].

4.3 Procedure

We used a pairwise comparison task in which two stimuli were presented sequentially, for 2 seconds each, with a 1 second inter-stimulus interval, and the observer was asked to select the one that looked more realistic. The exact wording on the briefing form was: "which of these scenes appears most similar to a real scene on the basis of your everyday viewing in a perceptive and cognitive sense". The short presentation time ensured that the observers were able to recall the first interval. The two stimulus intervals depicted the same scene, but under different conditions. They could be either a real scene shown in the illuminated box, or a virtual scene shown on our display with or without a distortion of varying magnitude (reduced resolution and contrast). Because we could not rapidly change the content of the real-scene box, the experiment was split into four sessions, one for each scene. The pairs to compare were selected using an active-sampling algorithm, ASAP [Mikhailiuk et al. 2021], which selects the pairs that maximize information gain. ASAP reduces the number of comparisons necessary to collect data and improves the sensitivity (accuracy) of the measurements. Each observer completed 108 comparisons per session: 2 distortion types × 3 luminance levels × 3 batches of ASAP, each consisting of 4 comparisons (5 distortion levels including "Real" -1) and an additional 3×12 comparisons between real and virtual (12 per each luminance level) for better evaluation. ASAP ensures that each condition is compared with another one at least once per batch. Comparisons from each luminance level were presented in groups within sessions to ensure observers were adapted to the luminance of the stimuli. Before each luminance group, a grey screen at the desired luminance level was shown for 30 seconds to allow observers to adjust to the average light level of the session. The ordering of the sessions was

randomized per observer and the ordering of the luminance-level groups within each session was randomized per session.

4.4 Results scaling

The pairwise comparison results across all observers were scaled under Thurstone's Case V model into continuous Just-Noticeable-Difference (JND) scores [Perez-Ortiz and Mantiuk 2017], using *pwcmp* software[3]. The JND scale is designed to capture inter-observer variability — the standard deviation of inter-observer variability is fixed (to 1.05) across the JND scale so that 1 JND corresponds to 75% of the population noticing the difference. Such a scale provides interpretable scores near and beyond the discrimination threshold. Confidence intervals were obtained by bootstrapping (5000 samples). No outliers were found in the data.

5 Results and discussion

Below, we discuss separately the effect of resolution, contrast, and luminance on realism, how well our display could reproduce real scenes, and the implications for display technologies.

5.1 Resolution

Our results, averaged across the four scenes, shown in Figure 5(a), indicate a strong increase in realism from 15 ppd up to 60 ppd, at which point the gains in realism start to diminish. The sensitivity to resolution is the highest at the highest luminance (70 cd/m^2) as indicated by the slope of the yellow curve in the figure. Our results align well with measurements from other studies. Figure 5(a) plots the resolution limit for an average observer found by Ashraf et al. [2025], for black text on a white background at 100 cd/m^2, as a dashed magenta line. They found the resolution limit to be at 96 ppd, which is between the two highest resolutions measured in our experiment, where the JND curve flattens. They could estimate the exact threshold because their measurement protocol allowed for continuous modulation of the display's resolution by changing the viewing distance. However, their data cannot be used to quantify the loss of realism below the threshold (because they measured classical detection/discrimination thresholds).

We also plot in Figure 5(a) the measurements of Masaoka et al. [2013] (in black). We rescaled their individual pairwise comparison data using the same procedure as for our experiment (Section 4.4) to ensure the same JND scale. Because they used a different digital filter, we had to adjust their ppd values by a factor of 0.82. This factor was determined as follows: to produce lower-resolution images, we used the Lanczos ($a = 3$) filter for downsampling and a box filter for upsampling. Because Masaoka et al. used a sinc^3 filter, their images were sharper. To match both measurements, we found how much the frequency of their images must be reduced so that both filters match at the modulation of 0.25 (dashed lines in Figure 4). The luminance of the white chart in their experiment was ~42 cd/m^2, which was between 7 cd/m^2 and 70 cd/m^2 luminance levels in our experiment.

While the trends of our measurements and theirs are comparable, their data indicate that, for the lower resolutions, ppd values approximately 29% higher than ours were required to achieve comparable realism scores. It is difficult to identify the cause of this discrepancy, as it could be due to different presentation time (2 s for us, 10 s for their protocol), differences between displays (LCD, vs. custom HDR display) or individual differences between observers. It is also possible that binocular disparity, present in our experiment and missing in theirs, provided additional cues to realism that masked some of the resolution distortions.

5.2 Contrast

Our results show that reductions in contrast significantly decrease perceived realism in a largely linear manner with the γ parameter of our operator. We do not observe "flattening" near the best virtual condition, as we saw for resolution. This finding is not expected, as the contrast sensitivity of the human visual system is non-linear, and it is relevant for content production because it should inform the choice of the cost function used in tone mapping.

5.3 Luminance

All measurements were repeated at three luminance levels, with a median luminance of 0.7, 7 or 70 cd/m^2. Our expectation was that sensitivity to distortions would be smaller at lower luminance levels, at which our contrast sensitivity is also lower. Indeed, for our resolution measurements, we observe the slowest decrease in realism with decreasing resolution at the lowest luminance of 0.7 cd/m^2, shown in Figure 5(a). However, almost the opposite trend could be observed for contrast in Figure 5(b), with the slowest decrease in realism with reducing contrast observed at the highest luminance of 70 cd/m^2. We cannot easily explain this result, but we hypothesize that this was due to increases in perceived realism when viewing content on a display that was brighter and had a far larger dynamic range than those typically found in commercial stereoscopic displays, such as VR headsets. It should be noted that each luminance condition contained a high-dynamic-range scene, which spanned a large range of luminance values, as shown in the histograms in Figure S.4 of the supplementary. For this reason, our experiment could not isolate the effect of luminance to a single luminance level. Instead, the results show that for realistic scenes with a range of luminance levels, the observers could make consistent judgements about realism even at very low luminance levels. Further discussion can be found in Section C of the supplementary.

5.4 Real vs. virtual

The drop in realism scores between the first two data points on each curve in Figure 5 — "Real" and "Best virtual" — illustrate the limitations of our display (and/or capture procedure and calibration) in presenting truly realistic stimuli. The percentages shown in plots (a) and (b) next to the "Best virtual" condition indicate the proportion of responses that correctly indicated the real scene, adjusted for random guessing (0% corresponds to a random guess). The percentages vary from 2% to 65.3%, indicating that observers were more often wrong than right about which stimulus was real when adjusted for chance, across luminance levels (Crystal: 23.3%, Figures: 17%, Wooden Pyramid: 26.6%). The results shown in Figure 5 are the results of JND scaling, which includes all comparisons, and the transitivity of comparisons assumed in the Thurstone

[3] *pwcmp* software: https://github.com/mantiuk/pwcmp

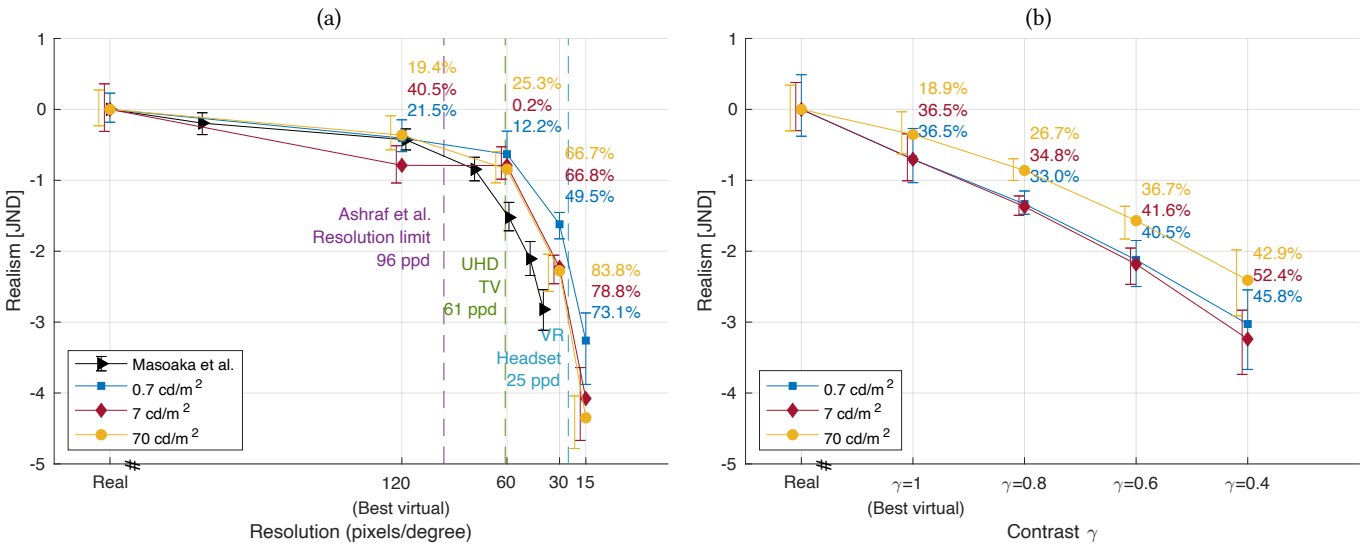


Fig. 5. Results scaled to JND values over all scenes. Different colors denote different average luminance levels for the stimuli. The error bars are shifted to avoid overlap. The percentages shown next to the markers indicate the likelihood that observers select the condition on the left of the marker as more realistic than the condition at the marker. The probabilities are adjusted for the guess rate, so that 0% means a random choice in a 2AFC experiment. The resolution plot also includes vertical lines indicating the ppd of a modern VR headset (Meta Quest 3) headset, a UHD TV when viewed at the recommended viewing distance (1.6 × display height), and the resolution limit for black text on a white background [Ashraf et al. 2025]

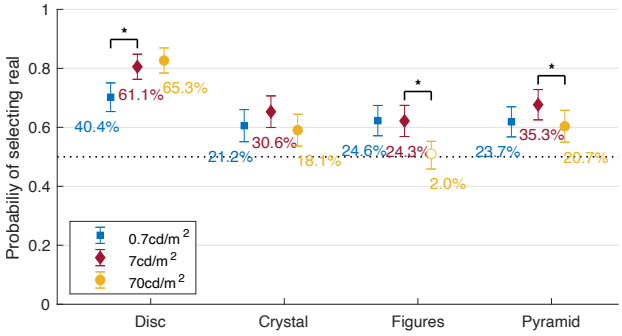


Fig. 6. The probability of correctly discriminating between the real and virtual scene, across four scenes and three luminance levels (colors). The chance rate was 0.5 for the 2AFC protocol. The error bars denote 95% confidence intervals. Unfilled markers indicate that there was no statistically significant difference from chance. The brackets with asterisks denote statistically significant differences between pairs of conditions shown at different luminance levels. The percentage numbers indicate the likelihood of selecting the real scene, when adjusted for chance.

Case V model. To better isolate real vs. virtual comparisons, we extracted only the comparisons between the real and best-virtual scenes from the measurements, for both resolution and contrast measurements, and plotted them per scene in Figure 6, adjusted for chance ($p' = 0.5 + 0.5\,p$).

Those results show that scene "Disc" was the most challenging to reproduce. It contained high-contrast text on a CD-ROM, and strong view-dependent effects due to the diffractive surface near the edge. For the other scenes, the probabilities varied between the guess rate (2%) and 35.3%. We observe a slight decrease in correct responses for the luminance of 0.7 cd/m^2, but the difference was only statistically significant for the "Disc" scene (one-tailed binomial test, $\alpha = 0.05$). Interestingly, we also observed a decrease in correct responses for the highest luminance of 70 cd/m^2, which, however, was statistically different only for the "Figures" and "Pyramid" scenes. This effect was small, and it could be attributed to the phenomenon discussed in Section 5.3 above — the brighter scenes looked unlike a standard dynamic range (SDR) display and, therefore, were perceived as more realistic, even though they were compared with the real scene of the same luminance. Overall, our display could reproduce realistic 3-D scenes with a very high level of realism, though small inaccuracies, in particular in color reproduction, could provide hints for selecting real objects. Color was particularly difficult to capture as our camera had different spectral sensitivity than the human eye. Because of that, it was impossible to find a bijective mapping between camera RGB values and cone responses [Finlayson et al. 2015]. This problem could potentially be avoided if the images were captured with an imaging colorimeter, as done in [Masaoka et al. 2013], or if we could capture multiple images, each with specifically optimized illuminance, to better infer colorimetric values [Gao et al. 2024].

We also note that some subtle visual cues, such as focus and parallax cues, were intentionally weak in the stimuli shown to observers. We achieved this through the design of the chosen scenes and by restricting head movement through the use of a forehead rest. While we did not study the impact of these cues, other work [Kim et al. 2024; March et al. 2022] indicates that they do aid in observer judgments of realism. We found no evidence of the effect of visual acuity when comparing observers with normal visual acuity (20/20;

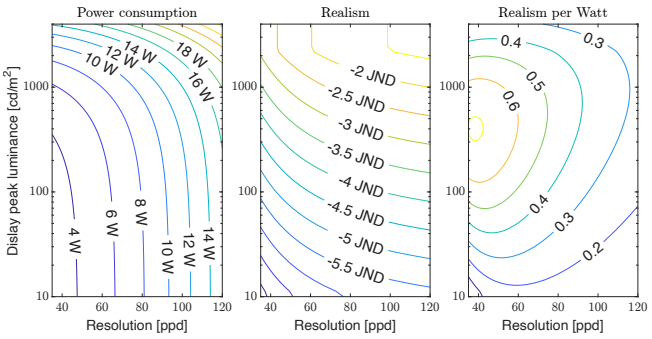


Fig. 7. Example of display realism vs. power analysis that can be performed using our measurements. We assume an AR display, with a field-of-view of 51.4°×55° and affected by 10 cd/m^2 of ambient light. Higher display peak luminance delivers higher dynamic range, but it requires more power. Higher resolution requires rendering more pixels, which also affects power consumption. Our measurements let us find the biggest gain in realism per watt of power spent. The reported values are provided as an illustrative example, based on multiple assumptions as stated in the text.

N=10) and observers with higher visual acuity (20/15 or 20/13; N=13) (p = 0.05, one-tailed, Bonferroni corrected).

5.5 Observer preconceived notions of realism

In our study, we relied on observers having a preconditioned understanding of what a 'real' scene looks like in order to perform the task as intended. While 'realism' is in itself a subjective quality, we anchor observer expectations of realism by including a real scene in our stimuli. We believe that it is reasonable to assume that observers can differentiate the experience of viewing lower quality displayed content (i.e., lower resolution, lower contrast, etc.) from that of viewing the real world. Our results show that observers consistently considered higher resolution stimuli to appear more real than lower resolution stimuli. This gives us confidence that observers performed the task as intended and that instances in which observers chose a virtual scene as appearing more real than a real scene indicate that the observer could not differentiate between the sensation of viewing the virtual scene and the real world.

5.6 Implications for future display systems

Our scale provides a useful guide for developers of future display technologies wishing to present perceptually realistic content, against which display capabilities can be measured. For example, while we do observe a plateau in increases in realism above resolutions of 60 ppd, consistent with the findings of Ashraf et al. [2025], we note that many contemporary commercial AR/VR displays have resolutions well below this level. Our measurements quantify the gain in realism achieved when increasing the resolution of such displays. Because our measurements express effects of degrading both resolution and contrast in the same JND units, we can consider both dimensions together when making decisions about display design. As an example, let us consider a hypothetical OST-AR headset, with a field of view of 51.4°×55°, and which is seen in a bright environment so that the displayed content is affected by 10 cd/m^2 of ambient light. With a few assumptions (Meta Quest Pro display power estimates from [Chen et al. 2024], 3 W GPU peak at 4.5 MPix, 50% of GPU power devoted to shading pixels), we can estimate power consumption for a range of resolutions and display peak luminance levels, as shown in the leftmost panel of Figure 7. Then, we made a further assumption that the displayed content spans 9 stops of dynamic range (512:1 contrast) and calculated the contrast compression required to reproduce that range. With that information, we could use our data to find the realism score for the combination of both parameters (assuming their independence), as plotted in the middle panel. Finally, we could estimate the gain in realism with a single watt of power, which we plot in the right panel. Such analysis indicates that for the current generation of AR displays, designed for bright environments, it is more beneficial to expend power increases on increasing peak brightness rather than resolution.

6 Conclusions

To make future AR/VR displays more realistic, and capable of blending real and virtual scenes, we will need to improve both their resolution and dynamic range. Our study measured the degradation in realism, scaled in JND units, due to insufficient resolution and contrast compression, where the latter is required to fit content into a limited display dynamic range. Importantly, the measurements were collected on an ultra-realistic 3-D display, capable of reproducing high dynamic range, high resolution, and geometric projection. This let us compare virtual with real scenes, and therefore anchor our measurements to the absolute scale of realism. Indeed, we demonstrated that for most tested scenes, observers were often unable to discriminate between the real and virtual objects. Furthermore, we repeated the measurements at three luminance levels, spanning three orders of magnitude, to capture the effect of display luminance.

Our resolution measurements matched the trend found in previous works [Masaoka et al. 2013], that gains in realism diminish rapidly after a threshold which is below the resolution limit of the the human visual system (96 ppd [Ashraf et al. 2025]). We note that the threshold at which improvements begin to diminish is lower when viewing content containing binocular disparity, as presented on our display, than when viewing content which does not contain binocular disparity, as in Masaoka et al. [2013]. It should be noted that the reported values are for an average observer, and the resolution threshold could be higher for a portion of the population with higher acuity. Our measurements showed that absolute luminance has an effect on the sensitivity to resolution, with the gains in realism climbing to 120 ppd for the brightest condition. Our contrast measurements showed an almost linear drop in realism with the reduction of contrast. However, unexpectedly, we saw the weakest reduction in realism due to contrast compression for the highest luminance condition. We speculate that brighter scenes inherently appear more realistic. Our display is still limited in many respects: we cannot reproduce parallax due to head motion, and the scenes must contain small objects on a black background, as we cannot reproduce continuous variations in focal depth. Despite these limitations, the stimuli remain far closer to real scenes than those used

in most studies. These restrictions are necessary to ensure the measurements reflect perceptual sensitivity rather than limitations of a particular display.

We hope that our measurements will help with better-informed decisions when designing future AR/VR displays. Our realism scale indicates how far we are from achieving the ultimate goal of reproducing reality and what changes in display capabilities are required to bring us closer to that goal.

7 Acknowledgments

We would like to thank Akshay Jindal, Jize Sha, and Ali Özgür Yöntem for their help with the display and in the early stages of the project. The display used in the experiments was constructed with the support of the European Union's Horizon 2020 research and innovation programme under the European Research Council (ERC) Consolidator Grant agreement N° 725253 (EyeCode) and under the Marie Skłodowska-Curie grant agreement N° 765911 (RealVision). We would also like to thank the anonymous reviewers for their valuable feedback and the experiment participants for their contribution to this research.

References

Maliha Ashraf, Alexandre Chapiro, and Rafał K. Mantiuk. 2025. Resolution limit of the eye — how many pixels can we see? *Nature Communications* 16, 1 (Oct. 2025), 9086. doi:10.1038/s41467-025-64679-2

M. Borg, S. S. Johansen, D. L. Thomsen, and M. Kraus. 2012. Practical Implementation of a Graphics Turing Test. In *Advances in Visual Computing*, George Bebis, Richard Boyle, Bahram Parvin, Darko Koracin, Charless Fowlkes, Sen Wang, Min-Hyung Choi, Stephan Mantler, Jürgen Schulze, Daniel Acevedo, Klaus Mueller, and Michael Papka (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 305–313.

Peter Burns and Don Williams. 2018. Camera Resolution and Distortion: Advanced Edge Fitting. *Electronic Imaging* 2018 (01 2018), 1–5. doi:10.2352/ISSN.2470-1173.2018.12.IQSP-171

Kenneth Chen, Thomas Wan, Nathan Matsuda, Ajit Ninan, Alexandre Chapiro, and Qi Sun. 2024. PEA-PODs: Perceptual Evaluation of Algorithms for Power Optimization in XR Displays. *ACM Transactions on Graphics* 43, 4 (2024), 1–17. doi:10.1145/3658126

David G. Curry, Gary L. Martinsen, and Darrel G. Hopper. 2003. Capability of the human visual system. In *Cockpit Displays X*, Darrel G. Hopper (Ed.), Vol. 5080. International Society for Optics and Photonics, SPIE, 58 – 69. doi:10.1117/12.502607

G. Eilertsen, R. K. Mantiuk, and J. Unger. 2017. A comparative review of tone-mapping algorithms for high dynamic range video. *Computer Graphics Forum* 36, 2 (May 2017), 565–592. doi:10.1111/cgf.13148

Graham D. Finlayson, Michal Mackiewicz, and Anya Hurlbert. 2015. Color Correction Using Root-Polynomial Regression. *IEEE Transactions on Image Processing* 24, 5 (May 2015), 1460–1470. doi:10.1109/TIP.2015.2405336

H. Gao, R. K. Mantiuk, and G. D. Finlayson. 2024. Color-Accurate Camera Capture with Multispectral Illumination and Multiple Exposures. *Computer Graphics Forum* 43, 7 (Oct. 2024), e15252. doi:10.1111/cgf.15252

Matthew Hamilton, Nicholas Wells, and Amilcar Soares. 2022. On Requirements for Field of Light Displays to Pass the Visual Turing Test. In *2022 IEEE International Symposium on Multimedia (ISM)*. 86–87. doi:10.1109/ISM55400.2022.00019

Param Hanji, Fangcheng Zhong, and Rafał K. Mantiuk. 2020. Noise-Aware Merging of High Dynamic Range Image Stacks Without Camera Calibration *(Lecture Notes in Computer Science, Vol. 12537)*. Springer International Publishing, Cham, 376–391. doi:10.1007/978-3-030-67070-2_23

Darrel G. Hopper. 2000. 1000 X difference between current displays and capability of human visual system: payoff potential for affordable defense systems. In *Cockpit Displays VII: Displays for Defense Applications*, Darrel G. Hopper (Ed.), Vol. 4022. International Society for Optics and Photonics, SPIE, 378 – 389. doi:10.1117/12.397759

Aaron Isaksen, Leonard McMillan, and Steven J. Gortler. 2000. Dynamically reparameterized light fields. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '00)*. ACM Press/Addison-Wesley Publishing Co., USA, 297–306. doi:10.1145/344779.344929

Dongyeon Kim, Seung-Woo Nam, Suyeon Choi, Jong-Mo Seo, Gordon Wetzstein, and Yoonchan Jeong. 2024. Holographic Parallax Improves 3D Perceptual Realism. *ACM Transactions on Graphics* 43, 4 (2024), 1–13. doi:10.1145/3658168

Rafał K. Mantiuk, Karol Myszkowski, and Hans-Peter Seidel. 2015. *High Dynamic Range Imaging* (1 ed.). Wiley, 1–42. doi:10.1002/047134608X.W8265

Joseph March, Anantha Krishnan, Simon Watt, Marek Wernikowski, Hongyun Gao, Ali Özgür Yöntem, and Rafal Mantiuk. 2022. Impact of correct and simulated focus cues on perceived realism. In *SIGGRAPH Asia 2022 Conference Papers*. Association for Computing Machinery, New York, NY, USA, Article 22, 9 pages. doi:10.1145/3550469.3555405

Joseph G. March, Anantha Krishnan, Rafał K. Mantiuk, and Simon J. Watt. 2024. Impact of focus cue presentation on perceived realism of 3-D scene structure: Implications for scene perception and for display technology. *Journal of Vision* 24, 2 (02 2024), 13–13. arXiv:https://arvojournals.org/arvo/content_public/journal/jov/938664/i1534-7362-24-2-13_1709035688.01908.pdf doi:10.1167/jov.24.2.13

Kenichiro Masaoka, Yukihiro Nishida, Masayuki Sugawara, Eisuke Nakasu, and Yuji Nojiri. 2013. Sensation of Realness From High-Resolution Images of Real Objects. *IEEE Transactions on Broadcasting* 59, 1 (2013), 72–83. doi:10.1109/TBC.2012.2232491

Nathan Matsuda, Alex Chapiro, Yang Zhao, Clinton Smith, Romain Bachy, and Douglas Lanman. 2022. Realistic Luminance in VR. In *SIGGRAPH Asia 2022 Conference Papers*. Association for Computing Machinery, New York, NY, USA, Article 21, 8 pages. doi:10.1145/3550469.3555427

Gary W. Meyer, Holly E. Rushmeier, Michael F. Cohen, Donald P. Greenberg, and Kenneth E. Torrance. 1986. An experimental evaluation of computer graphics imagery. *ACM Trans. Graph.* 5, 1 (jan 1986), 30–50. doi:10.1145/7529.7920

Aliaksei Mikhailiuk, Clifford Wilmot, Maria Perez-Ortiz, Dingcheng Yue, and Rafal Mantiuk. 2021. Active Sampling for Pairwise Comparisons via Approximate Message Passing and Information Gain Maximization. In *2020 IEEE International Conference on Pattern Recognition (ICPR)*. doi:10.1109/ICPR48806.2021.9412676

Maria Perez-Ortiz and Rafał K. Mantiuk. 2017. A practical guide and software for analysing pairwise comparison experiments. arXiv:1712.03686 (Dec. 2017). http://arxiv.org/abs/1712.03686 arXiv:1712.03686 [cs, stat].

Erik Reinhard, Wolfgang Heidrich, Paul Debevec, Sumanta Pattanaik, Greg Ward, and Karol Myszkowski. 2010. *High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting*. Morgan Kaufmann. http://books.google.com/books?id=w1i_1kejoYcC&pgis=1

Helge Seetzen, Wolfgang Heidrich, Wolfgang Stuerzlinger, Greg Ward, Lorne Whitehead, Matthew Trentacoste, Abhijeet Ghosh, and Andrejs Vorozcovs. 2004. High dynamic range display systems. *ACM Transactions on Graphics* 23, 3 (Aug. 2004), 760. doi:10.1145/1015706.1015797

Jialin Wang, Rongkai Shi, Xiaodong Li, Yushi Wei, and Hai-Ning Liang. 2024. Omnidirectional Virtual Visual Acuity: A User-Centric Visual Clarity Metric for Virtual Reality Head-Mounted Displays and Environments. *IEEE Transactions on Visualization and Computer Graphics* 30, 5 (2024), 2033–2043. doi:10.1109/TVCG.2024.3372127

Marek Wernikowski, Joseph G. March, Radosław Mantiuk, Ali Özgür Yöntem, and Rafał K. Mantiuk. 2024. Tracking Eye Position and Gaze Direction in Near-Eye Volumetric Displays. In *2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 787–796. doi:10.1109/ISMAR62088.2024.00094

Fangcheng Zhong, Akshay Jindal, Ali Özgür Yöntem, Param Hanji, Simon J. Watt, and Rafał K. Mantiuk. 2021. Reproducing reality with a high-dynamic-range multi-focal stereo display. *ACM Trans. Graph.* 40, 6, Article 241 (dec 2021). doi:10.1145/3478513.3480513

Supplementary Document: Quantifying reality of ultra-realistic 3-D displays — the effect of resolution and contrast

A Observer characteristics

25 observers took part in our experiment. Figure S.1 shows the histogram of their ages, and Figure S.2 shows the histogram of their visual acuity. We note that the observers were primarily in the 15–35 age range, at which visual acuity typically peaks.

Six observers wore corrective glasses during the experiment, all of whom were myopic without glasses. As the display consists only of transmissive and reflective optical components, we do not expect wearing glasses to significantly impact the results.

Eleven observers reported some prior experience with rendered virtual content, such as video games viewed either on a traditional 'flat' display or on a VR headset.

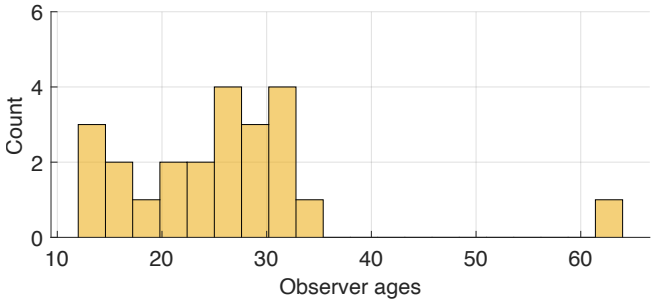


Fig. S.1. Histogram of observer ages. Observers from a wide range of age groups participated in the experiment, including one observer who was 63 year old. Note that the age distribution is primarily in the 15–35 year old range.

B Per scene results

Figure S.3 shows the results of the experiment plotted separately for each scene. The format of the results shown follows that of the results presented in the main paper. The y-axis has been extended to accommodate the larger error bars present in the individual scene measurements due to the lower number of comparisons. We note that all scene measurements have the same overall shape, although the effect of luminance and distortions differs somewhat between the scenes.

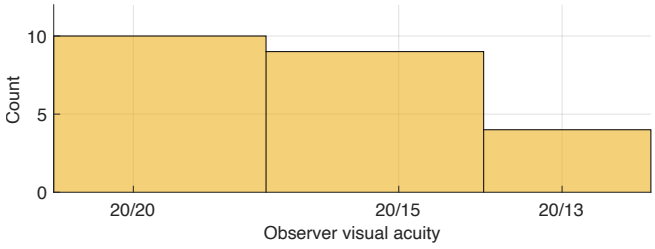


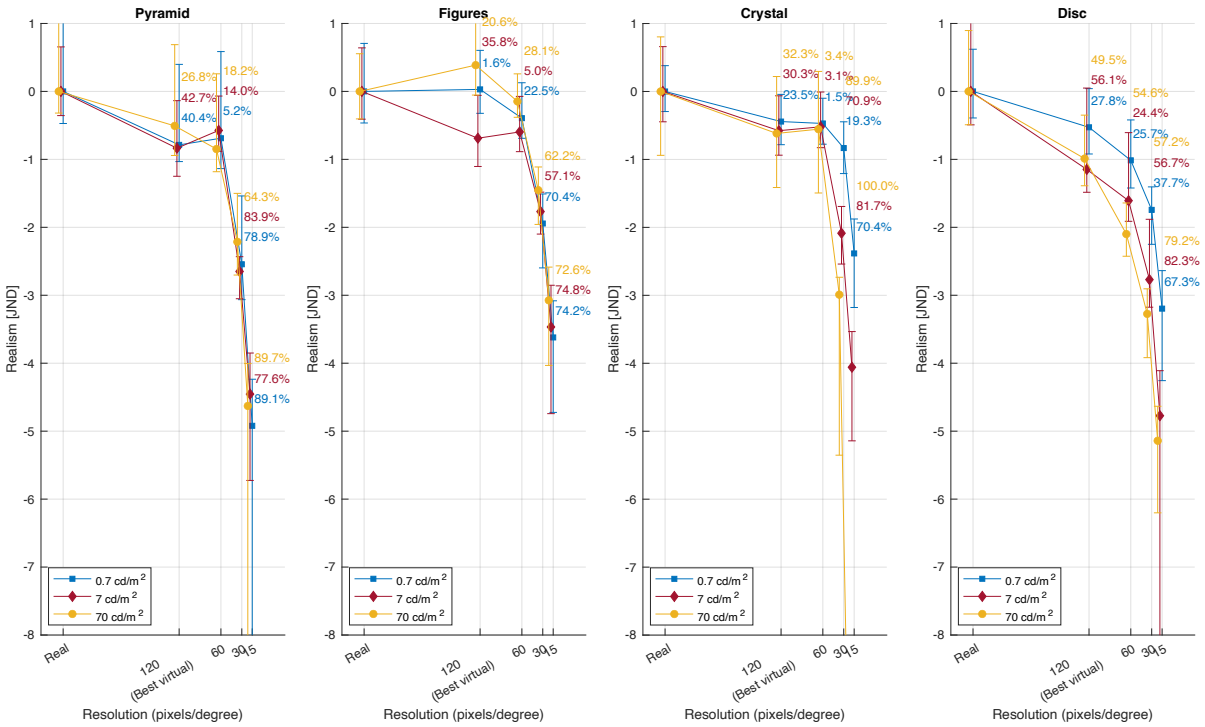
Fig. S.2. Histogram of observer visual acuity scores.

C Effect of luminanance

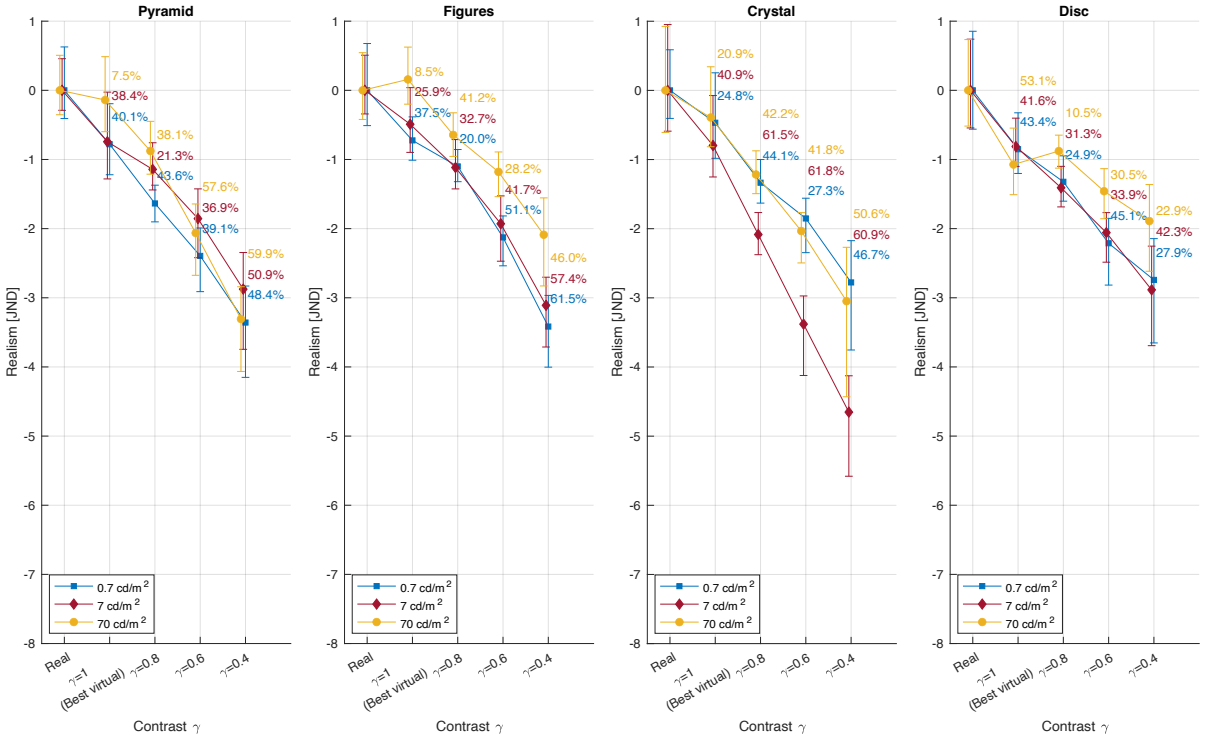
Our results, shown in Figure 5, do not indicate a consistent reduction in sensitivity as the scene's luminance decreases. The drop of realism due to resolution is the slowest for the lowest scene luminance (0.7, blue line), but it is the slowest for the highest luminance (70 cd/m^2, orange line) when contrast is reduced. This result may indicate that near-threshold sensitivity does not translate to suprathreshold perception of realism. This is not completely unexpected, as many suprathreshold phenomena are unaffected by thresholds (e.g., contrast constancy). It should also be noted that the scenes spanned a range of luminance for each mean luminance (see Figure S.4), and the judgements were likely based on the most visible (the brightest) part of the scene.

D Effect of contrast adjustment

Figure S.4 shows visual effect (top row) the image histograms (bottom row) of our 'Crystal' scene at each contrast value and luminance level.



(a) Resolution



(b) Contrast

Fig. S.3. Results scaled to JND values, plotted separately for each scene. Different colors denote different average luminance levels for the stimuli. The error bars are shifted to avoid overlap. The percentages shown next to the markers indicate the likelihood that observers select the condition on the left of the marker as more realistic than the condition at the marker. The probabilities are adjusted for the guess rate, so that 0% means a random choice in a 2AFC experiment.

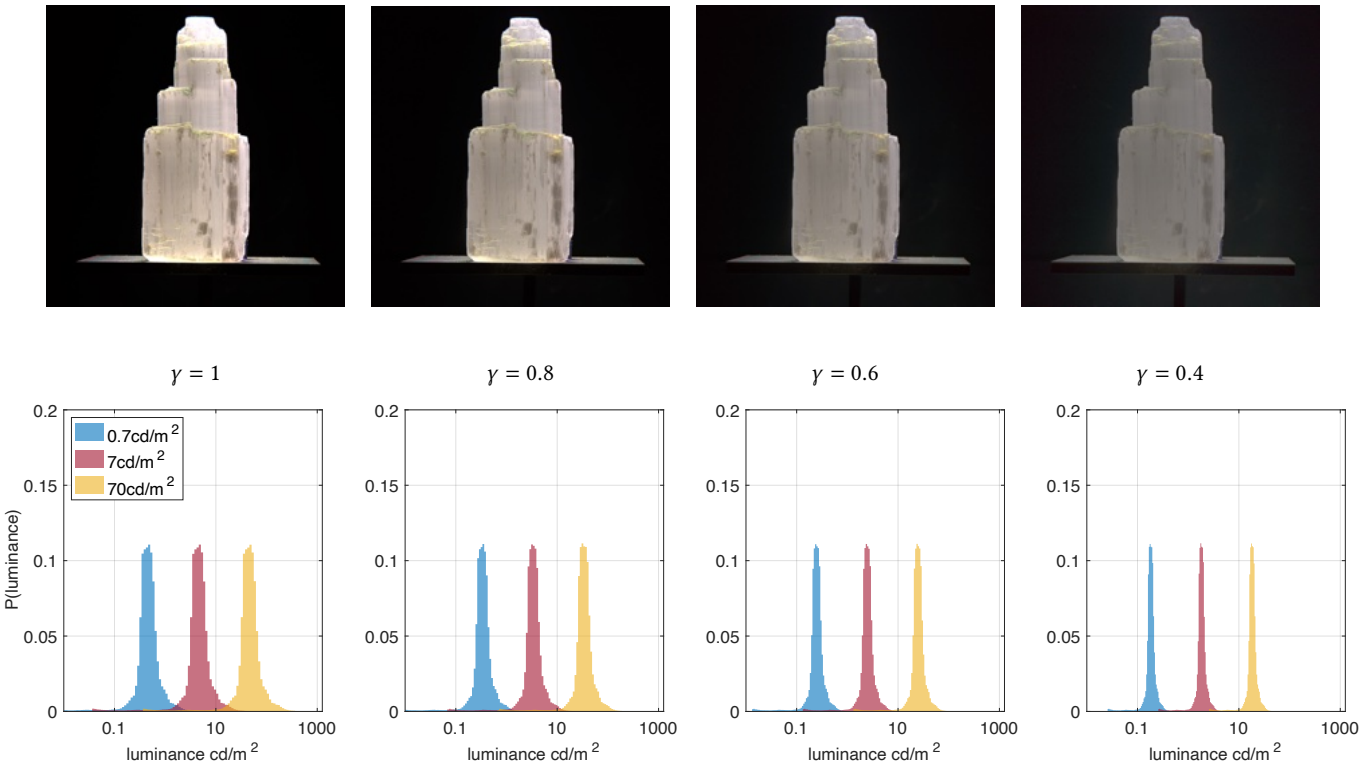


Fig. S.4. Illustration of the visual impact of contrast reduction on the 'crystal' scene. Examples of the rendered output at each alpha level in the 0.07 cd/m^{2}illumination condition are shown in the top row. Luminance histograms for the crystal scene in all three illumination conditions (0.7 cd/m^2, 7 cd/m^2, 70 cd/m^{2}from the top) and at all chosen γ values are shown in subsequent rows.