Sensation of Realness From High-Resolution Images of Real Objects

Kenichiro Masaoka, *Member, IEEE,* Yukihiro Nishida, Masayuki Sugawara, *Member, IEEE,*
Eisuke Nakasu, *Member, IEEE,* and Yuji Nojiri

Abstract—We performed subjective assessments to quantify the sensation of realness for images at various angular resolutions and that of their real-object counterparts using a paired-comparison procedure. Both the images and real objects were viewed through a synopter, which removed horizontal disparity and presented the same images to both eyes. The size, perspective, luminance, and chromaticity of the images were reproduced to be identical to those of the real objects. Eighty-two observers with normal vision were asked to choose the viewed image that appeared most similar to the real object for each pair of images. The results indicated that the realness of images increased steadily as the image resolution increased up to around 60 cycles per degree, whereafter it gradually approached that of the real objects.

Index Terms—Angular resolution, Bradley–Terry paired-comparison model, MTF, synopter, visual realness.

I. INTRODUCTION

$\mathbf{I}$N EVERYDAY life, we can easily distinguish display images from real objects by cues such as display chassis, image size and perspective, luminance and chromaticity, or binocular disparity and motion parallax. We can do so even if these factors are carefully controlled to be identical between images and their real objects, as images have a pixel structure whereas real objects do not. That is, the resolution of an image limits its state of appearing real in terms of visual perception or "visual realness."

The ability to resolve spatial detail at a certain distance is defined as visual acuity. A visual angle of 1 arcminute gap in an optotype is frequently considered to be the normal value of resolution in humans. Regarding the requirements for the subjective assessment of television image quality, it is often reasoned that if the pixel size is smaller than the separable angle, the scanning line or pixel structure becomes invisible. The ITU-R recommends a "design viewing distance" [1] of 3H for HDTV [2] and 6H for SDTV [3] (H: screen height). The image resolution at these viewing distances is

TABLE I
VISUAL ACUITY NOTATIONS OF 20/20 OR BETTER

Snellen Equivalent			Visual Angle	
Feet	Meters	Decimal	MAR (min.)	LogMAR
20/20	6/6	1.0	1.0	0.0
20/16	6/4.8	1.25	0.8	−0.1
20/12.5	6/3.8	1.6	0.63	−0.2
20/10	6/3	2.0	0.5	−0.3
20/8	6/2.4	2.5	0.4	−0.4

about 30 cpd (cycles per degree) or one pixel per arc minute, which corresponds to the separable visual angle at a normal visual acuity of 20/20 in Snellen notation. The numerator and denominator of the Snellen fraction represent the maximum legible viewing distance in feet for the observer and a person with normal visual acuity, respectively, using a Snellen eye chart.

The average visual acuity of healthy eyes is better than 20/20. In fact, the cone photoreceptor spacing limits the visual acuity to between −0.4 and −0.3 logMAR (MAR: minimum angle of resolution) [4]. Velasco e Cruz [5] traced the historical roots of 20/20 as an incorrect standard of normal visual acuity. Hermann von Helmholtz (1821–1894) reported that the resolution of normal vision was approximately 1 arcminute "per line pair," which corresponds to 20/10. However, Herman Snellen (1834–1908) mistook the resolution for 1 arcminute "per line" when he built his well-known visual acuity chart. As a result, many visual acuity charts for values above 20/20 have been truncated on the general assumption of Snellen's normal visual acuity.

There are several notations to record visual acuity. Table I shows corresponding notations of visual acuities for 20/20 or better. A geometric or logarithmic progression in the size of optotypes, conforming to the Weber-Fechner law, provides a linear increase in perception. The logarithmic representation of the MAR, introduced by Bailey and Lovie [6], approximates an equal discriminability scale for visual acuity [7] and is superior in its scientific principles. The linearity of the logMAR scale allows for calculating the proper average visual acuity by taking the arithmetic mean of the logMAR values [8], [9].

Elliott *et al.* [10] and Frisén and Frisén [11] measured visual acuity changes throughout adulthood in normal, healthy eyes. The measurements were taken with logMAR charts from subjects with optimal refractive correction. The mean best-corrected visual acuity peaks at better than −0.15 logMAR in

the mid to late 20s, whereafter it gradually declines with age in approaching 0.0 logMAR. The supposed standard of 20/20 for visual acuity may be said to represent the lower limit of normal. In fact, normal visual acuity is recommended as a screening cutoff for the subjective assessment of the quality of television pictures [12].

When we consider visual acuities better than 20/20, an image resolution of 30 cpd is insufficient to assure visual realness of images. Moreover, we should consider the spatial frequency characteristics of images in their entirety rather than the angular resolution corresponding to the viewer's MAR alone. From our perspective, the sensation of realness of images should steadily increase as the image resolution increases above 30 cpd, and thereafter gradually reach "the visual realness of the real object" at some image resolution. Before quantitatively evaluating the sensation of realness, it is important to contemplate the optimal spatial frequency characteristics.

The modulation transfer function (MTF) is the magnitude response of the optical system to sinusoids of different spatial frequencies. However, the sampled imaging system is not shift-invariant; that is, the modulation of a sinusoid depends on the phase with respect to the sampling sites of the imager. This shift variance violates one of the main assumptions required for a convolution analysis of the image-forming process. To preserve the convenience of a transfer-function approach, we assume that the object being captured has spatial frequency components whose phases are random and uniformly distributed with respect to the sampling sites. Thus, the sampling MTF is represented by the ensemble average of all possible MTFs and assumed to have a shift-invariant quantity [13].

For a square sampling grid, the sampling impulse response is a two-dimensional rectangular function whose sides are all equal to the sampling interval. If each detector integrates the scene irradiance over its square surface (with each side equal in length to the sampling interval), i.e., in contiguous detectors, the detector-footprint MTF can be written as

$$\text{MTF}_{\text{detector}}\left(f_x, f_y\right) = \left|\text{sinc}\left(f_x, f_y\right)\right| \tag{1}$$

where f_x and f_y are the horizontal and vertical spatial frequencies, respectively, normalized by the sampling frequency. Koren [14] approximated the characteristic overall MTF of typical dedicated film scanners and digital camera sensors as sinc^3 on the basis of measured data.

The purpose of this study was to estimate the angular resolution required for the visual realness of images. We compared real objects and their faithfully reproduced images displayed on a medical LCD, whereby the luminance and chromaticity distributions of the real objects were virtually within the dynamic range and color gamut of the display.

II. METHODS

A. Observers

Eighty-two naïve observers (39 males and 43 females) in their 20 s (mean 26.0 y, S.D. 3.5) participated in the experiments. All observers had normal vision for their left and right eyes wearing their own habitual correction if required (naked: 27 observers, glasses: 32 observers, contacts: 23 observers). Their binocular visual acuities ranged from 0.0 to −0.4 logMAR. (1, 22, 42, 15, and 2 observers with 0.0, −0.1, −0.2, −0.3, and −0.4 logMAR, respectively, mean −0.19 logMAR, S.D. 0.078).

Each observer's visual acuity was measured using Tumbling E optotypes [15] because they can be rendered as bi-level images. Landolt Cs or other optotypes with round shapes cannot be rendered as bi-level images because they would contain jagged edges. Single black Es were displayed in the center of a 21.3-inch 2560 × 2048 pixel TFT monochrome LCD, which was mounted on a linear stage moving in the 4274 to 4538 mm range of the viewing distance to adjust the visual angle of the gap of each logMAR optotype (8, 6, 5, 4, and 3-pixel gaps for 0.0, −0.1, −0.2, −0.3, and −0.4 logMAR, respectively). The observer was stabilized in a chin- and headrest with his/her habitual correction. The background white level was 200 cd/m^2 and the Michelson contrast was greater than 99% under 130 lux illumination. Single Es ranging from 0.0 to −0.4 logMAR in 0.1 logMAR steps were presented twenty times to each observer with the criterion of 62.5% correct for 4AFC (alternative forced choice). A warning sound preceded stimulus presentation by 0.1 s. The observer was asked to identify the orientation of the E and make his/her best guess when uncertain. No feedback was given.

All observers' refractive errors, which were less than ±6.00 DS or 2.50 DC, were measured with a built-in infrared optometer (AA-2000 Accommodometer, NIDEK Co., Ltd., Gamagori, Japan). Each observer's interpupillary distance (IPD) was also measured with the optometer. The observers had passed a stereo vision test (Stereo Fly SO-001; Stereo optical Co., Inc., Chicago, IL, USA) and the Ishihara color test.

B. Apparatus

Fig. 1 shows a diagram of the apparatus layout and pictures of the setup. The position and alignment of the apparatus was adjusted with a laser distance meter and a cross-line self-leveling laser level. A black tunnel and a darkroom were used to block out any stray light. Many heavy weights were used to stabilize the apparatus.

1) Illuminator: Real objects were illuminated with two simulated daylight lamps (SOLAX XC-100A; Seric Ltd., Tokyo, Japan) placed on the left and right sides in front of the objects. The spectral distribution of the light was rendered similar to that of sunlight using xenon technology and spectral correction filters. A light shaping diffuser film with elliptical output of 1° (horizontal) × 60° (vertical) in FWHM (LSD60×1PC10-F5; Luminit LLC, Torrance, CA, USA) was set in front of each lamp to improve vertical uniformity.

2) Imaging Colorimeter: Real objects were shot with a CCD-based imaging colorimeter (PM-1400; Radiant Imaging, Inc., Redmond, WA, USA) and a telephoto lens with a fixed focal length of 300 mm (Ai AF-S Nikkor 300 mm f/4D IF-ED; Nikon Corp., Tokyo, Japan) focused at f/16. The colorimeter had a 3072 × 2048 pixel (pixel size: 9 × 9 μm) 2-stage Peltier-cooled full-frame CCD with a 14-bit A/D converter. The colorimeter was equipped with a motorized internal filter-
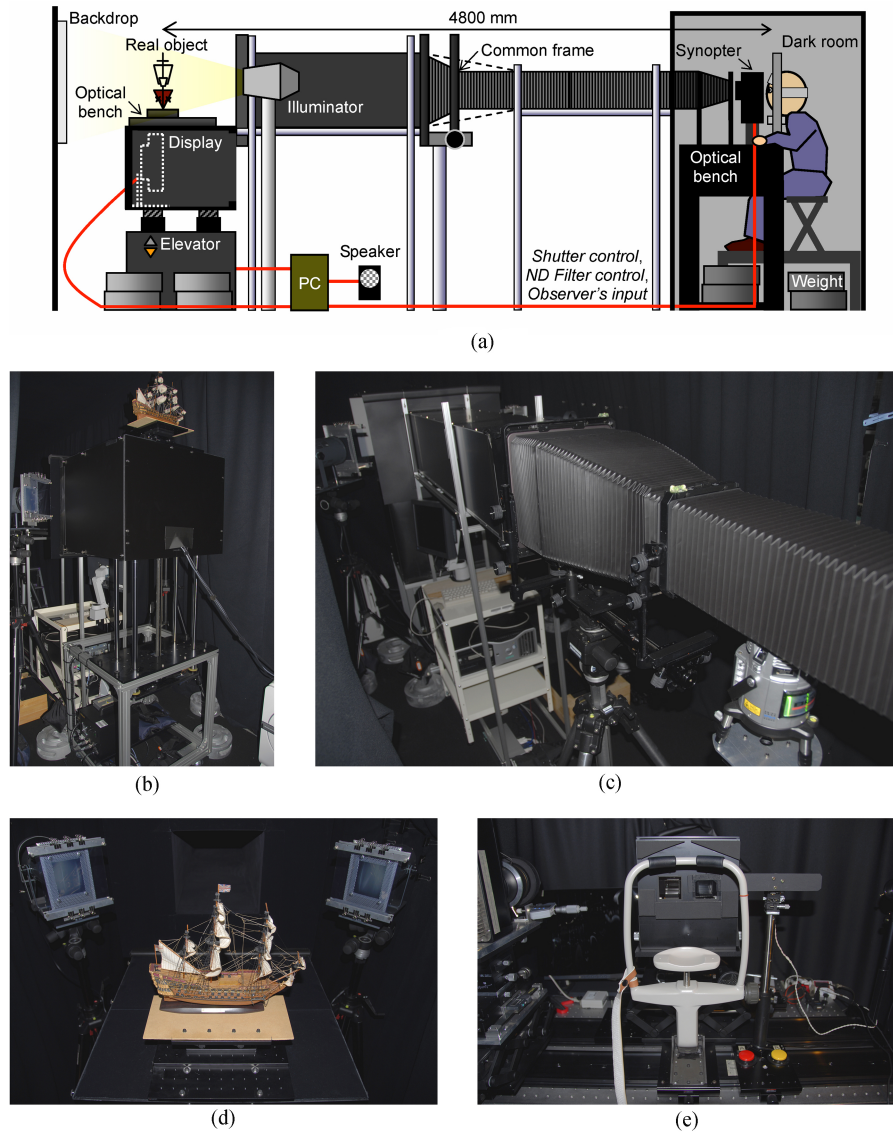
Fig. 1. Apparatus layout. (a) Diagram of the entire apparatus and pictures of the (b) elevator, (c) black tunnel, (d) illuminators with a real object on the optical bench, and (e) inside of the dark room. When shooting, the common frame was not installed with the prismoid bellow extended, as shown by the broken lines in (a).

wheel housing four color filters: X, Y, Z, and XB. The spectral sensitivities of the Y and Z filters were designed and fabricated to closely approximate the CIE 1931 $\bar{y}$ and $\bar{z}$ color matching functions, respectively. The X and XB filter shapes closely approximated the longer and shorter humps of the $\bar{x}$ function, respectively. The shooting distance (from the first principal point of the lens to the precise focus position) was 4800 mm. The colorimeter was set at the position of the observer in Fig. 1 without the synopter when shooting. One hundred twelve frames were shot for each filter in alternation and averaged to reduce the photon shot noise. A dark image (shutter closed) was also taken every 8 frames, and the averaged dark image was subtracted from the averaged filter images.

3) *Display:* For displaying images, a TFT color LCD monitor (RadiForce RX220; Eizo Nanao Corp., Hakusan, Japan) was used. The screen had an area of 1600×1200 pixels (pixel size: 0.270×0.270 mm), and was calibrated to have uniform color and brightness with a gamma of 2.4 across the entire screen. The display supported 10-bit input for each RGB color via a DisplayPort connection. The monitor's 12-bit look-up table had a palette of 4096 gray scales and outputted the most appropriate 1024 grayscale tones (10 bits) for red, green, and blue.

4) *Synopter:* The displayed images and real objects were viewed through a synopter, which was a simple optical device consisting of one beam splitter and three mirrors. The synopter presented identical images to the observers' left and right eyes without horizontal disparity. Fig. 2 shows a schematic diagram and picture of the synopter. The left and right mirrors were adjusted separately to be positioned just in front of each observer's left and right eyes on the basis of his/her IPD. A light balancing filter (Fuji filter LBA-1; Fujifilm Corp., Tokyo, Japan) and two neutral density filters (Fuji filter ND-0.2; Fujifilm Corp., Tokyo, Japan) were used to correct the spectral distribution. The light balancing filter and one of the neutral density filters were fixed to the left and right outputs of the
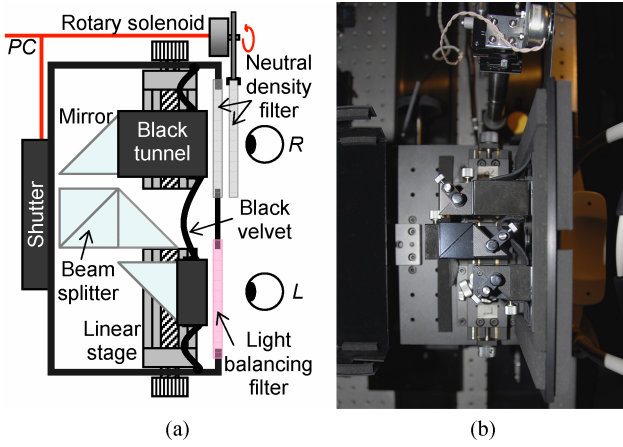
Fig. 2. Synopter system. (a) Diagram and (b) picture of the synopter system viewed from above with the front cover opened.
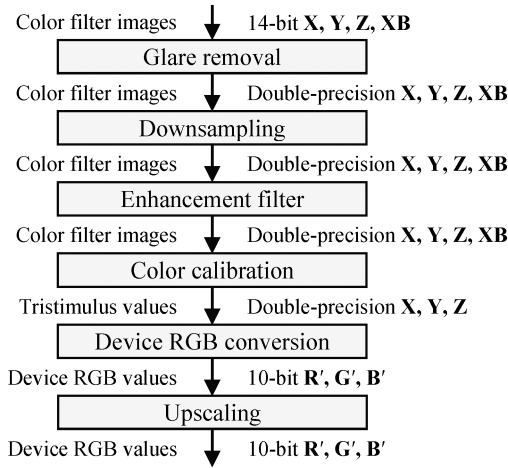

Fig. 3. Image processing flow chart.

TABLE II
PARAMETER VALUES FOR IMAGE PROCESSING

Resolution (cpd)	Low-Pass Filter Size	Downscaling Ratio	Upscaling Ratio
155	5×5	1/2	1
78	11×11	1/4	2
53	17×17	1/6	3
39	21×21	1/8	4
31	27×27	1/10	5
26	31×31	1/12	6

analytically, some simplifications are made. We used a deconvolution method for glare removal, assuming an empirical Gaussian-like glare function h_{glare} for each color filter:

$$h_{\mathrm{glare}}(x, y) = k \exp\left(-\left(x^2 + m^2 y^2\right)^{p/2} \big/ 2\sigma\right) \qquad (2)$$

where x and y were the pixel coordinates with $(x, y) = (0, 0)$ at the center of h_{glare}, $-1024 \le x$, $y \le 1023$. To estimate the parameters k, m, p, and σ in (2), a black anodized aluminum plate with a hole backlit with an integral sphere (LE6-100; Image Engineering GmbH & Co. KG, Frechen, Germany) was set at the precise focus position and shot with the colorimeter. The radius of the hole was about 110 pixels on the image sensor. The glare component $I_{\mathrm{glare}}(x, y)$ was estimated from the shot image $I_{\mathrm{shot}}(x, y)$, $-1024 \le x$, $y \le 1023$, with a central area of 2048×2048 pixels, as follows:

$$I_{\mathrm{glare}} = I_{\mathrm{shot}} \otimes h_{\mathrm{glare}}. \qquad (3)$$

Here, $\otimes$ denotes the two-dimensional convolution operation. The glare-corrected image I_{corr} was written as

$$I_{\mathrm{corr}} = \left(I_{\mathrm{shot}} - I_{\mathrm{glare}}\right) \big/ (1 - K) \qquad (4)$$

where $K = \sum_{x,y} h_{\mathrm{glare}}(x, y)$. The parameters were optimized iteratively for each color filter image so that the glare around the hole was removed by minimizing G:

$$G = \sum_{150 < r < 750} I_{\mathrm{corr}}(r)^2 \big/ (1 - K)^2 \qquad (5)$$

where $r = \sqrt{x^2 + y^2}$.

Fig. 4 shows the normalized luminance distribution of the target image for each color filter before and after glare removal. The glare distribution around the hole differed across color filters. After glare removal, the luminance level around the hole was reduced to approximately less than 0.1% of that of the hole. Although some pixels had negative values because of the simplicity of the glare removal model, most of the absolute values were also less than 0.1% of that of the hole.

Fig. 5 shows the spatial frequency response (SFR) of each color filter of the imaging colorimeter before and after glare removal. The SFR for each color filter was measured on the basis of slanted-edge analysis with the SFR algorithm [17] using a knife-edge target of a black anodized aluminum plate backlit with the integral sphere. As can be seen, the low frequency cusp behavior due to glare [18] was subdued while retaining the overall MTF intrinsic to the shooting system (assuming no veiling glare). The multi-directional SFRs [19] were also measured using a starburst target of a black anodized

synopter system, respectively; the other neutral density filter was inserted by a rotary solenoid when viewing the LCD to correct the transmission difference due to polarization of the LCD screen. The synopter had a shutter system (SSH-R; Sigma Koki Co., Ltd., Tokyo, Japan) in front of the beam splitter.

C. Image Processing

The color filter images (X, Y, Z, and XB data) shot with the colorimeter were processed to have 6 different angular resolutions ranging from 26 to 155 cpd with accurate color reproduction. For each image resolution, sinc^3 was used as the optimal overall MTF. This subsection describes each image-processing step following the flow chart shown in Fig. 3. Table II shows the parameter values.

1) *Glare Removal:* The image plane of an optical or electro-optical imaging system normally receives not only image-forming radiation; it also receives stray radiation, which can reduce image contrast. This unwanted radiation, known as veiling glare, usually arises from one or more causes, such as scattering in lens elements, reflections off the lens barrel and/or surfaces, and reflections from the sensor surface [16]. Given the difficulty or impossibility of deriving the glare components
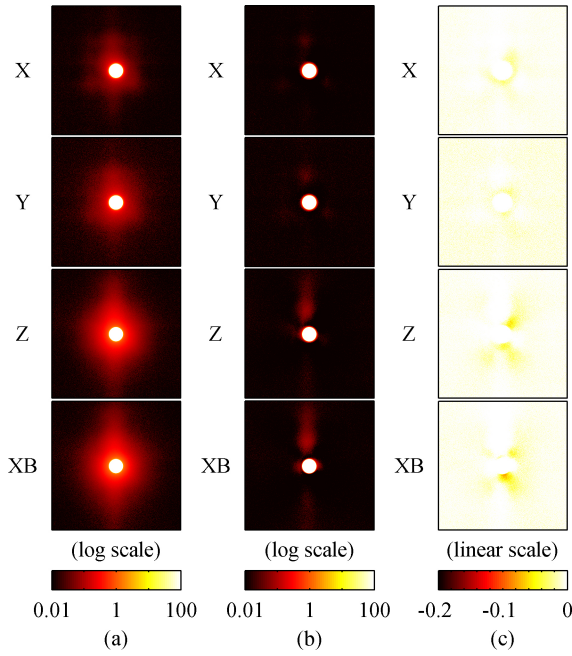
Fig. 4. Luminance distribution of the target for each color filter before and after glare removal. The X, Y, Z, and XB filter images are shown from top to bottom. (a) Left column shows the luminance distributions before glare removal, and (b) middle and (c) right columns show the luminance distributions after glare removal in positive and negative values, respectively.
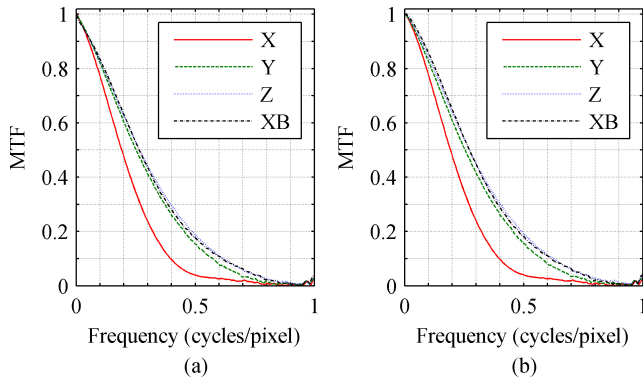

Fig. 5. SFR of each color filter: (a) before and (b) after glare removal.

aluminum plate backlit with the integral sphere. This revealed that the SFRs were approximately identical, i.e., the overall MTF was circularly symmetric.

2) *Downsampling:* The glare-corrected images (X, Y, Z, and XB data) were filtered by an anti-aliasing low-pass filter for each image resolution. The low-pass filters were two-dimensional FIR filters designed by the window method [20] using a Hamming window. The filter size depends on the destination resolution; larger filter sizes were required for lower destination resolutions, as shown in Table II. The $((n + 1)/2,\ (n + 1)/2)$-th entry of the $n \times n$ matrix containing the desired frequency response was 1, whereas other entries were zero. The low-pass filtered data were downscaled into six different image sizes. The downscaling ratios are shown in Table II.

3) *Enhancement Filter:* The downscaled images were filtered by unsharp contrast enhancement filters so that the MTFs
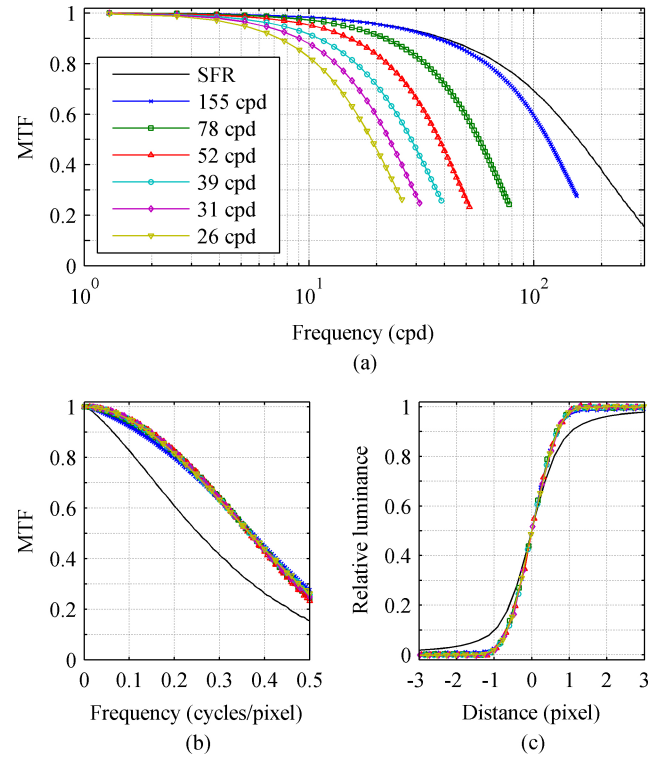

Fig. 6. Spatial frequency characteristics of the Y filter. (a) MTFs as a function of frequency in cpd. (b) MTFs as a function of frequency in cycles/pixel. (c) Edge profiles.

simulated $sinc^3$. Two-dimensional (3×3) negative Laplacian filters were used for unsharp contrast enhancement. The filter h_{en} was written as

$$h_{en}=(1 - s) \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \frac{s}{\alpha + 1} \begin{bmatrix} -\alpha & \alpha - 1 & -\alpha \\ \alpha - 1 & \alpha + 5 & \alpha - 1 \\ -\alpha & \alpha - 1 & -\alpha \end{bmatrix} \quad (6)$$

where s was optimized iteratively for each color filter and destination image resolution, and α was set to 0.36 to make the MTFs at lower image resolutions circularly symmetric.

Fig. 6(a) shows the MTFs of the Y-filter image after enhancement for the six destination angular resolutions. The horizontal axis is in cycles per degree. Fig. 6(b) shows the same MTFs with spatial frequencies normalized by the sampling frequency, i.e., in cycles/pixel. The MTFs were similar and approximated $sinc^3$. The MTFs of the other color filters were also close to $sinc^3$. Fig. 6(c) shows the edge profile of the Y filter at each destination angular resolution. There is virtually no over-shooting or under-shooting of the edges at all angular resolutions. The same was true for the other filters.

4) *Color Calibration:* The X, Y, Z, and XB data were transformed to tristimulus XYZ values using a 3×4 color correction matrix. To estimate the matrix, a color calibration target (ColorChecker Classic; X-Rite Inc., Grand Rapids, Michigan, USA) was shot with the colorimeter under the same lighting condition. For references, the tristimulus XYZ values of the 24 color patches were measured with a spectroradiometer (SR-UL1; Topcon Technohouse Corp., Tokyo, Japan). The

TABLE III

MEAN AND STANDARD DEVIATION OF COLOR ERRORS (CIE94) OF THE 24 COLOR PATCHES FOR THE CALIBRATED COLORIMETER

	Color Filters	
Glare	X, Y, Z	X, Y, Z, XB
Not corrected	1.0318 ± 0.7387	0.6320 ± 0.3165
Corrected	0.9369 ± 0.7617	0.5539 ± 0.3586

TABLE IV

LUMINANCE LEVEL AND CORRELATED COLOR TEMPERATURE OF THE REAL WHITE CHART AND THE IMAGE DISPLAYED THROUGH THE SYNOPTER

	Left	Right
Real	42.4 cd/m^2, 6138 K	41.9 cd/m^2, 6091 K
Display	41.6 cd/m^2, 6098 K	42.8 cd/m^2, 5976 K

coefficients of the color matrix were optimized by minimizing the sum of the squares of the color errors (CIE94 [21]).

Table III shows the mean color errors and their standard deviations before and after glare correction with three (X, Y, and Z) and four color filters (X, Y, Z, and XB). The advantages of removing glare and using the XB color filter were validated on the basis of the accuracy of color reproduction.

5) *Device RGB Conversion:* The tristimulus XYZ values of each image were subtracted by the XYZ values of the black level of the display and converted to the linear device RGB (red-green-blue) values using a 3×3 color conversion matrix on the basis of the luminance and chromaticity of the RGB primaries:

$$\begin{bmatrix} R_{x,y} \\ G_{x,y} \\ B_{x,y} \end{bmatrix} = \mathbf{M} \begin{bmatrix} X_{x,y} - X_{\text{black}} \\ Y_{x,y} - Y_{\text{black}} \\ Z_{x,y} - Z_{\text{black}} \end{bmatrix}. \tag{7}$$

The linear RGB values were clipped to the range $[0, 1]$ and encoded with the inverse of the display's electro-optical conversion function (EOCF, or "gamma") of 2.4, forming 10-bit non-linear R$'$G$'$B$'$ values.

6) *Upscaling:* Finally, the R$'$G$'$B$'$ images were scaled up to 1024×1024 pixels using the nearest-neighbor method to simulate a square-pixel structure. The upscaling ratios are shown in Table II. The 1024×1024 pixel images were trimmed to the central area of 1000×1000 pixels when viewing with a black frame installed in the black tunnel shown in Fig. 1.

D. Luminance Level Adjustments

The output levels of the illuminators and display were adjusted so that the luminance and chromaticity became identical through the synopter.

Fig. 7 shows the spectral radiance distribution of an illuminated white balance target (ColorChecker White Balance; X-Rite Inc., Grand Rapids, Michigan, USA) and the reproduced image through the synopter. The luminance levels and correlated color temperatures are shown in Table IV. The display's contrast ratio was 690:1.

Fig. 8 shows the human visual modulation threshold and the modulations of the quantization error in the ramp and
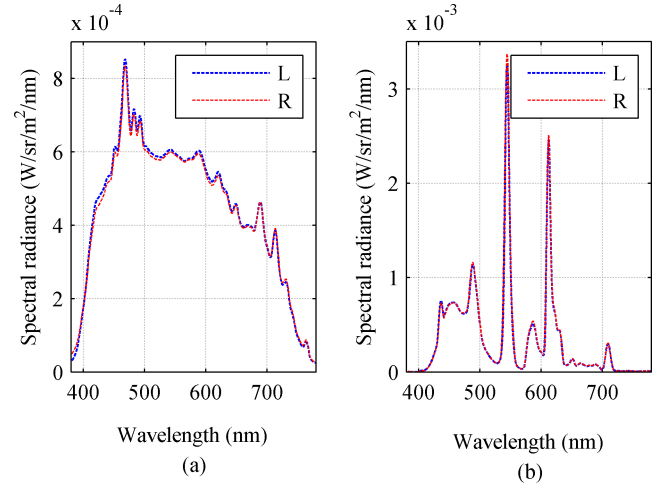

Fig. 7. Spectral radiance distributions of the left and right outputs of the synopter for (a) illuminated white chart and (b) image displayed on the LCD screen. The two lines in each plot considerably overlap each other.
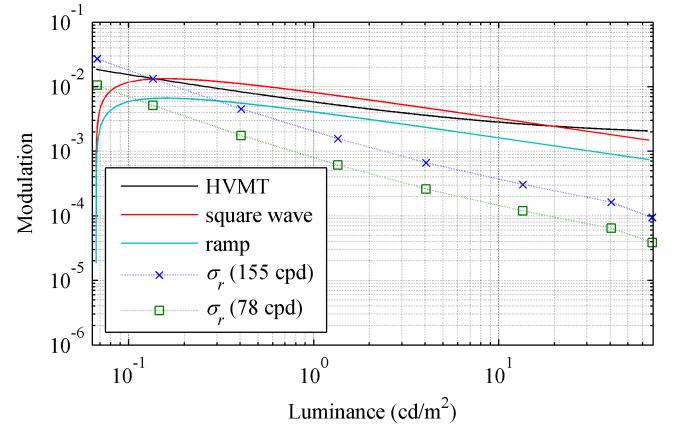

Fig. 8. Human visual modulation threshold compared with the modulations of errors in square wave and ramp patterns reproduced by the 10-bit display with a gamma of 2.4, and the relative standard deviations of the luminance for the Y-filter images at 155 and 78 cpd.

square wave stimulus rendered with the 10-bit display with a gamma of 2.4. The threshold was calculated on the basis of Barten's model [22] at the adaptation luminance levels within the display's luminance range with a field of view (FOV) of $3.2° \times 3.2°$. The contrast sensitivity in humans depends on spatial frequency. The threshold in Fig. 8 indicates the highest sensitivity at each adaptation luminance level, i.e., modulations under the threshold curve are invisible regardless of the spatial frequency. The modulation of the quantization error was that of the fundamental frequency of the difference between the ideal ramp or square wave and the quantized stepwise waveform [23]. On the basis of the simulated results, any ramp waveform can be reproduced without contouring artifacts. Although some square waveforms are reproduced with visible artifacts, such patterns are rare and probably negligible in natural images. Fig. 8 also shows the relative standard deviations of the luminance σ_r for the Y-filter images of the uniform white target produced by the integral sphere at several luminance levels. The image resolutions of the Y-filter images were 155 and 78 cpd after glare removal, downsampling, and enhancement.
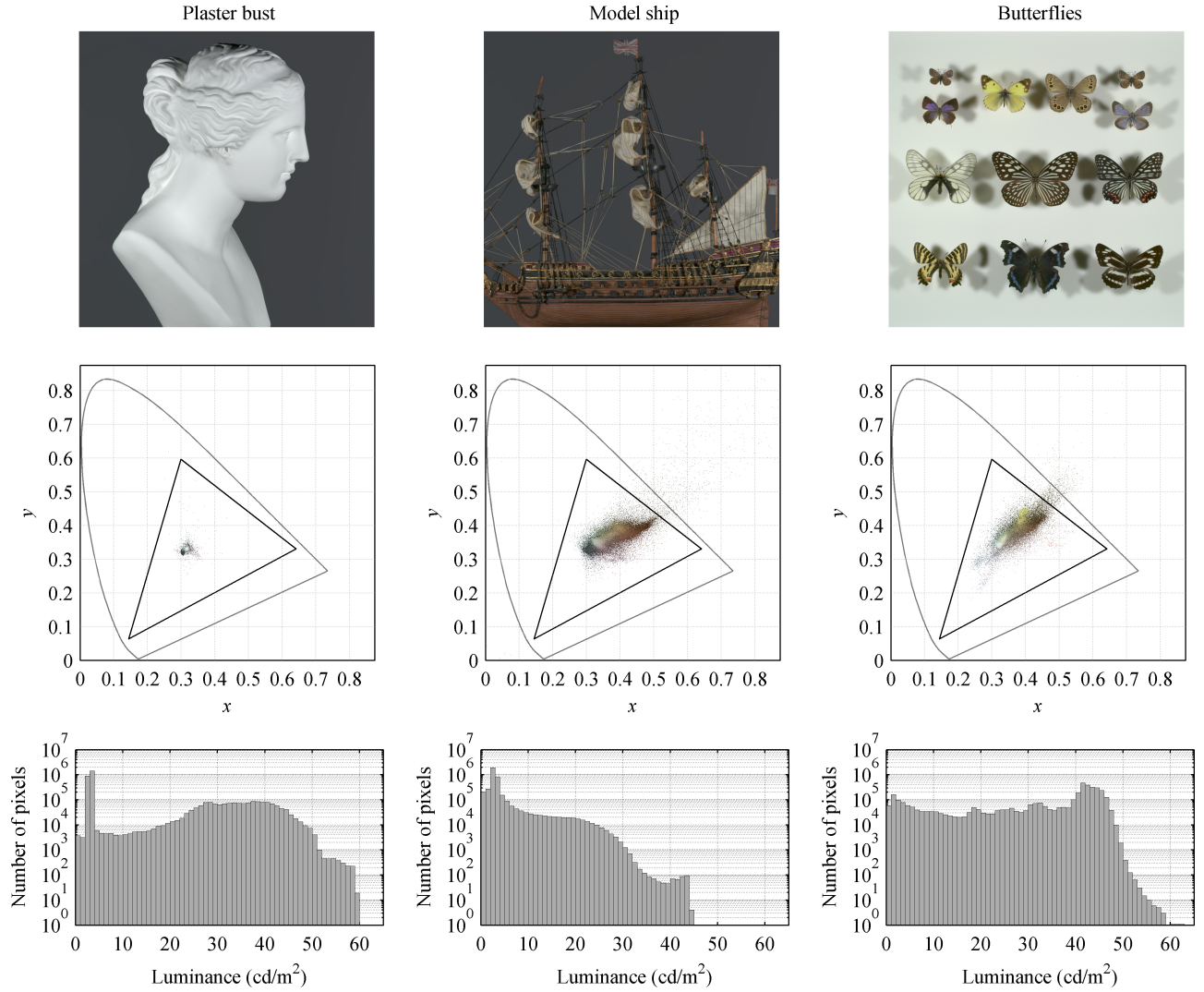
Plaster bust   Model ship   Butterflies

Fig. 9. Test images, chromaticity distributions (2000 $\times$ 2000 pixels), and luminance histograms (2000 $\times$ 2000 pixels).

E. Real Objects and Test Images

The display should have luminance and chromaticity ranges that reproduce those of real objects as accurately as possible to evaluate the effect of image resolution on visual realness. In addition, the depths of the objects should be within the depth of field of the colorimeter.

We carefully selected three objects, whose images are shown in Fig. 9. The "Plaster bust", which is often used as an artist's model, contains a sharp outline, waves in the hair, and subtle gradations across the surface. The "Model ship" has sails of fine texture, thin threads, and delicate parts, all of which are high spatial frequency components. The "Butterflies" are Japanese specimens whose wings have high contrast patterns with many detailed shapes and colors. Fig. 9 also shows the chromaticity distribution and luminance histogram of the 2000 $\times$ 2000 pixel image for each object after glare removal and color calibration. Some dark yellow pixels were out of the display's gamut because of over-reduction of the glare correction in the Z-filter images. The luminance levels of the histograms were calculated on the basis of the luminance level of the real white target through the synopter shown in Table IV. The

numbers of under-clipped pixels were 217, 6457, and 4850 for 1000 $\times$ 1000 pixel images (155 cpd) of the Plaster bust, Model ship, and Butterflies. No pixels were over-clipped. Fig. 10 shows the SFRs at the shooting distances of 4750, 4800 (precise focus position), and 4850 mm, as well as pictures of the Plaster bust and Model ship shown with ±50 mm scale bars.

F. Experiment

A paired-comparison procedure was used to quantify the sensation of realness of real objects and their images. Each observer had three sessions; one for each object. The session order was random among the observers. In each session, twenty-one pairs ($_7C_2$) of one real object plus the images of six different angular resolutions shown in Table II were presented, following a trial session with four or more pairs including a pair of images at 155 and 19 cpd (the latter resolution was used only for the trial session). Each observer was asked to choose the viewed image ("A" or "B") that appeared most similar to the real object on the basis of his/her everyday viewing in a perceptive and cognitive sense. The first stimulus "A" was presented for 10 s. Subsequently, the shutter of the
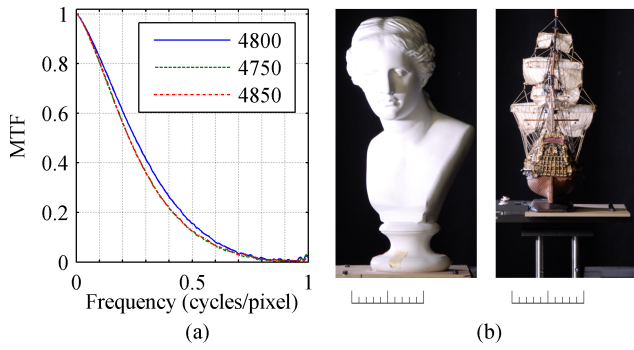
Fig. 10. Depth of field. (a) SFRs of the colorimeter at the viewing distances of 4800 mm and 4800 mm $\pm$ 50 mm. (b) Pictures of the plaster bust and model ship with $\pm$50-mm scale bars.

synopter was closed for 4 s. The second stimulus "B" was then presented until the observer made a choice with the buttons or 10 s had elapsed, after which the shutter was closed again and the next stimulus "A" was presented after 4 s. Each choice latency was recorded. The within- and between-pair orders were random. The order number was announced by a speaker before presenting each pair. Each presentation was preceded by voice "A" or voice "B". One session, including the trial session, required about 10 min. A 15-min break was given between sessions.

The observer was stabilized in a chin- and headrest with a head band in the dark room, as shown in Fig. 1. The viewing distance was 4800 mm, which was equal to the shooting distance. This allowed the perceived size and perspective of the image to appear identical to those of the real objects. A common frame for both the real objects and images was installed in the tunnel. The FOV was $3.2° \times 3.2°$. The observer viewed stimuli through the synopter and the tunnel. Prior to presenting the images, the elevator was raised to the upper position and a box housing the display was attached to the tunnel, completely intercepting the light from the illuminators. Next, the images were presented. When required, the display and the real object were exchanged; otherwise a sound to simulate the mechanical noise of the elevator was reproduced by the speaker so that the observer was unable to notice the up/down movement.

III. RESULTS

During trial sessions, all observers noticed differences in the angular resolutions between 155 and 19 cpd. Most observers reported differences in sharpness. Some reported differences in image contrast and/or three-dimensional appearance. No observer noticed that real objects were included in the stimuli, or reported any heterogeneities such as in luminance or chromaticity during any of the sessions. The average choice latency for every combination of pair and object was less than 4 s.

Table V shows the frequency matrix for each object. According to goodness-of-fit tests, the test statistic (chi-square with 15 d.f.) revealed that the Bradley–Terry model (see Appendix) fit the choice data well for each object: Plaster

bust, $\chi_0^2 = 13.25$, $p = 0.58$, n.s.; Model ship, $\chi_0^2 = 9.76$, $p = 0.83$, n.s.; and Butterflies, $\chi_0^2 = 21.91$, $p = 0.11$, n.s. The logarithms of the estimates for the true ratings $\ln p_i$ with the 95% confidence interval (CI) are shown in Fig. 11. No significant difference between the two gender groups regardless of the object was found (chi-square with 6 d.f.): Plaster bust, $\chi_0^2 = 7.10$, $p = 0.31$, n.s.; Model ship, $\chi_0^2 = 3.61$, $p = 0.73$, n.s.; and Butterflies, $\chi_0^2 = 4.16$, $p = 0.65$, n.s.

Comparisons were made between three groups of subjects differing in binocular visual acuity to examine the effect of visual acuity: group Normal (23 observers with 0.0 or -0.1 logMAR), group Better (42 observers with -0.2 logMAR), and group Best (17 observers with -0.3 or -0.4 logMAR). The test statistic (chi-square with 15 d.f.) for goodness-of-fit found the Bradley-Terry model to fit the choice data well for each object and visual acuity group; Plaster bust, $\chi_0^2 = 11.94$, $p = 0.68$, n.s. for group Normal, $\chi_0^2 = 8.22$, $p = 0.91$, n.s. for group Better, $\chi_0^2 = 15.74$, $p = 0.40$, n.s. for group Best; Model ship, $\chi_0^2 = 11.98$, $p = 0.68$, n.s. for group Normal, $\chi_0^2 = 8.34$, $p = 0.91$, n.s. for group Better, $\chi_0^2 = 21.07$, $p = 0.13$, n.s. for group Best; and Butterflies, $\chi_0^2 = 22.14$, $p = 0.10$, n.s. for group Normal, $\chi_0^2 = 17.22$, $p = 0.31$, n.s. for group Better, $\chi_0^2 = 6.16$, $p = 0.98$, n.s. for group Best. The test statistic (chi-square with 12 d.f.) for treatment selection equality of these three visual acuity groups showed a significant difference for the Plaster bust ($\chi_0^2 = 59.8$, $p < 0.01$), Model ship ($\chi_0^2 = 36.9$, $p < 0.01$), and Butterflies ($\chi_0^2 = 36.5$, $p < 0.01$). The logarithms of the estimates of the true ratings $\ln p_i$ with 95% CI for each object and visual acuity group are shown in Fig. 12.

IV. DISCUSSION

The observers' binocular visual acuities may be said to represent that of normal, healthy eyes in the 20 s with habitual correction. The mean value (-0.19 logMAR) is better than the measured data in the mid to late 20 s (-0.15 logMAR) because of the screening cut-off of 0.0 logMAR.

The trend for the true ratings in Fig. 11 is clear; as image resolution increases, the sensation of realness steadily rises up to around 60 cpd, beyond which the increment becomes moderate in approaching that of the real object. The sensation of realness of images at around 30 cpd was significantly lower than that of the real object. The scores for the Plaster bust images at around 30 cpd were higher than those for the other objects. This could have occurred because the former had fewer high spatial-frequency components than the latter. The low scores of the Butterflies image at around 30 cpd would then have been due to the brighter background as well as the high-contrast components at high frequencies. The sensations of realness of images at 155 cpd were not significantly lower than those for the real objects. However, given the functional relationship, an image resolution of several hundred cpd would be required for perfect visual realness.

From Fig. 12, it is found that the scores at around 30 cpd depend on visual acuity; observers with higher visual acuities show a tendency of relatively lower sensation of realness. On the other hand, the image resolution over which the increments
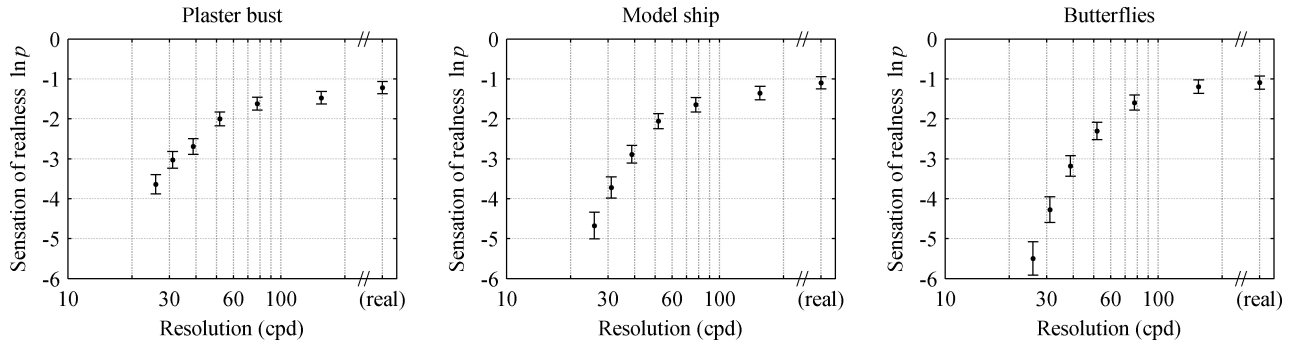
Fig. 11. Scores for the sensation of realness (the logarithms of the estimates for the true ratings) with the 95% CI for each object ($N = 82$).

TABLE V
FREQUENCY MATRICES

Plaster Bust								Model Ship								Butterflies							
(cpd)	real	155	78	53	39	31	26	(cpd)	real	155	78	53	39	31	26	(cpd)	real	155	78	53	39	31	26
real	—	48	46	56	68	69	77	real	—	44	48	63	73	78	78	real	—	47	43	67	74	78	81
155	34	—	48	56	61	63	74	155	38	—	48	53	69	73	78	155	35	—	48	70	72	76	80
78	36	34	—	52	63	68	66	78	34	34	—	50	61	72	78	78	39	34	—	51	63	76	81
53	26	26	30	—	59	62	70	53	19	29	32	—	57	71	77	53	15	12	31	—	62	76	79
39	14	21	19	23	—	48	64	39	9	13	21	25	—	61	68	39	8	10	19	20	—	63	73
31	13	19	14	20	34	—	51	31	4	9	10	11	21	—	64	31	4	6	6	6	19	—	65
26	5	8	16	12	18	31	—	26	4	4	4	5	14	18	—	26	1	2	1	3	9	17	—

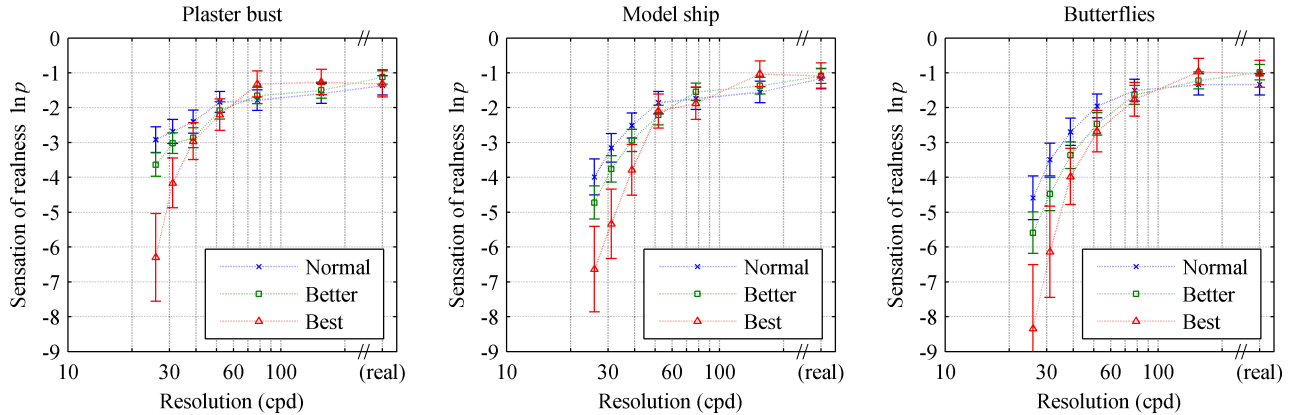Note: Row stimuli are chosen over column stimuli.

Fig. 12. Scores for the sensation of realness (the logarithms of the estimates for the true ratings) with the 95% CI for each object and visual acuity group ($N_{Normal} = 23$, $N_{Better} = 42$, $N_{Best} = 17$).

of the scores become moderate is around 60 cpd regardless of visual acuity. This is contrary to the expectation that the angular resolution depends on the observer's MAR. From an engineering perspective, an angular resolution of 60 cpd is necessary for the visual realness of images. Moreover, the observers in the experiments viewed images with a narrow FOV under dark surround. Given that the contrast sensitivity improves when viewing brighter or wider FOV images, a higher angular resolution would be necessary.

As it is obvious from Fig. 7, there were significant differences in the spectral radiations between the real objects and the displayed images. The differences might have caused observer metamerism and potentially biased the experimental results. Observer metamerism is an unavoidable fundamental limitation on colorimetry as well as on the experiment. However, the magnitude of mean-level observer metamerism is not significant (around 1 CIELAB unit when using trichromatic displays with 2-degree FOV) [24]. No observer reported any differences in luminance or chromaticity between the real objects and images. Analyzing the frequency matrices with the real-object elements excluded shows the same trend again, i.e., realness steadily increases up to 60 cpd, after which the increase becomes moderate. From these facts, it is rational to assume that observer metamerism was negligible and did not bias the experimental results.

In addition, the use of a synopter could have biased the experimental results. It is well known in the psychological literature that viewing through a synopter can lead to apparent depth or paradoxical stereopsis through accentuation of "pictorial relief" in images [25]. However, the observers thought they were viewing images rather than real objects in the experiments. It is interesting to observe that when

viewing extremely high angular-resolution images in normal viewing condition (i.e., even binocularly), the images look nicely rounded and solid, which implies that higher angular resolution as well as synoptic viewing enhances pictorial relief.

In actual display-viewing conditions, the angular image resolution is generally less than 60 cpd. In HDTV production, the details are usually enhanced to compensate for this low resolution. Thus, TV engineers often enhance picture sharpness by boosting the overall MTF beyond sinc^3. In the case of down-converting TV programs from HDTV to SDTV, a steep low-pass filter is typically used. In these situations, there is a trade-off between image sharpness and artifacts such as ringing and clipping [26]. In this sense, it is hoped that a higher image resolution will be realized for future video systems.

The results suggest a viewing distance of 6H or longer, rather than 3 H, is required for full HDTV resolution for visual realness. In the same way, 3 H or longer is required for 4 K resolution (3840×2160 pixels) and 1.5 H or longer for 8 K resolution (7680×4320 pixels) at the cost of sense of presence induced by FOV; the sense of presence of large-display images increases as the visual angle increases, reaching a plateau at around 80-degree horizontal FOV [27].

V. CONCLUSION

We performed subjective assessments to examine the sensation of realness of high-resolution images, comparing the images with real objects. We used a paired-comparison procedure to quantify the sensation of realness of images and their real-object counterparts. We found that the sensation of realness steadily increased above an image resolution of 30 cpd and lost momentum around 60 cpd, regardless of visual acuity differences in observers with normal vision. For visual realness, a resolution of 60 cpd or higher appears necessary. We hope that the data on image resolution for visual realness covered in this paper may serve as an impetus for further research into high-resolution cameras and displays.

APPENDIX: UNDERSTANDING THE BRADLEY-TERRY MODEL FOR PAIRED COMPARISONS

The method of paired comparisons is widely employed when subjective differences are small and fine judgments are required. In this method, objects or "treatments" are compared pairwise and judged subjectively. Thurstone first introduced the law of comparative judgment [28] and his Case V is frequently used as a probabilistic model for paired comparisons. An alternative model is the Bradley-Terry model [29], which assumes that each treatment has a true rating and that when compared with another treatment, the choice probability is determined by the ratio of the true ratings of the two treatments. Although Thurstone's Case V model is privileged within the imaging community, it is universally acknowledged that both models give very similar results in applications (e.g., [30], [31]). Moreover, the Bradley-Terry model is more mathematically developed [30] and gives more robust estimates of confidence intervals than Thurstone's Case V model [31].

Thurstone's Case V and Bradley-Terry models are both linear models of paired comparisons. The basic paired comparisons experiment has t treatments, $T_1, \ldots, T_t$, and n_{ij} comparisons of T_i with T_j, $n_{ji} = n_{ij}$, $i \neq j$, $i, j = 1, \ldots, t$. The number of times T_i is selected over T_j after n_{ij} comparisons is a_{ij}, and $a_{ij} = n_{ij} - a_{ji}$. David [32] supposes that T_i has "merit" V_i when judged on some characteristic. In a linear (paired-comparison) model, these merits can be represented on a linear scale, and the probability π_{ij} of choosing T_i over T_j is

$$\pi_{ij} = H(V_i - V_j) \qquad (8)$$

where H is a symmetric distribution function, $H(-x) = 1 - H(x)$, $H(-\infty) = 0$, $H(\infty) = 1$.

In the Bradley-Terry model, a simple logistic function,

$$H(x) = 1/(1 + e^{-x}) \qquad (9)$$

is used for H, instead of the normal cdf with zero mean and an arbitrary value of the standard deviation used in Thurstone's Case V model. Equation (8) can be rewritten as

$$\ln(\pi_{ij}/(1 - \pi_{ij})) = V_i - V_j. \qquad (10)$$

It is postulated that the t treatments have true ratings, $\pi_1, \ldots, \pi_t$, on a particular subjective continuum throughout an experiment. These ratings represent relative selection probabilities for the treatments and the continuum is specialized by the requirements that $\pi_i \geq 0$ and $\Sigma \pi_i = 1$, the latter being added for specificity. Further definition follows by assuming that

$$\pi_{ij} = \pi_i / (\pi_i + \pi_j) \qquad (11)$$

so that (10) simply takes the form $\ln \pi_i - \ln \pi_j = V_i - V_j$ and the scale parameters

$$V_i = \ln \pi_i \qquad (12)$$

are provided.

On the assumption of independent responses for the comparisons, the likelihood function L is

$$L = \prod_i \pi_i^{a_i} / \prod_{i<j} (\pi_i + \pi_j)^{n_{ij}} \qquad (13)$$

where $a_i = \sum_{j, j \neq i} a_{ij}$. The estimates of the true ratings of t treatments, $p_1, \ldots, p_t$, are solved iteratively by maximizing ln L. A small MATLAB program is presented below. In terms of calculation of confidence intervals for the estimated parameters and chi-square tests for selection equality, differences between groups, and goodness-of-fit, we refer the reader elsewhere (e.g., [29], [30]).

```
function p = btmodel(M)
% M: t x t paired comparison matrix
% p: estimates of the true ratings of the t treatments
t = length(M);
p = fminsearch(@(p) -btlik(p, M), ones(1, t-1)/t);
p = [p 1-sum(p)];

function lnL = btlik(p, M)
p = [p 1-sum(p)];
if any(p<=0), lnL = -Inf; return, end
```

```
t = length(p);
lnL = 0;
for i = 1:t-1
    j = i+1:t;
    pi_ij = p(i)./(p(i) + p(j));
    lnL = lnL + sum(M(i, j).*log(pi_ij) + M(j, i)'.*log (1-pi_ij));
end
```

REFERENCES

[1] *Relative Quality Requirements of Television Broadcast Systems*, ITU-R Rec. BT 1127, 1994.

[2] *Subjective Assessment Methods for Image Quality in High-Definition Television*, ITU-R Rec. BT 710-4, 1998.

[3] *Subjective Assessment of Standard Definition Digital Television (SDTV) Systems*, ITU-R Rec. BT 1129-2, 1998.

[4] R. A. Applegate, "Limits to vision: Can we do better than nature?" *J. Refract. Surg.*, vol. 16, no. 5, pp. S547–S551, Sep.–Oct. 2000.

[5] A. A. Velasco e Cruz, "Historical roots of 20/20 as a (wrong) standard value of normal visual acuity," *Optom. Vis. Sci.*, vol. 67, no. 8, p. 661, Aug. 1990.

[6] I. L. Bailey and J. E. Lovie, "New design principles for visual acuity letter charts," *Am. J. Optom. Physiol. Opt.*, vol. 53, no. 11, pp. 740–745, Nov. 1976.

[7] G. Westheimer, "Scaling of visual acuity measurements," *Arch. Ophthalmol.*, vol. 97, no. 2, pp. 327–330, Feb. 1979.

[8] M. J. Moseley and H. S. Jones, "Visual acuity: Calculating appropriate averages," *Acta Ophthalmologica*, vol. 71, no. 3, pp. 296–300, Jun. 1993.

[9] J. T. Holladay, "Proper method for calculating average visual acuity," *J. Refract. Surg.*, vol. 13, no. 4, pp. 388–391, Jul.–Aug. 1997.

[10] D. B. Elliott, K. C. Yang, and D. Whitaker, "Visual acuity changes throughout adulthood in normal, healthy eyes: Seeing beyond 6/6," *Optom. Vis. Sci.*, vol. 72, no. 3, pp. 186–191, Mar. 1995.

[11] L. Frisén and M. Frisén, "How good is normal visual acuity? A study of letter acuity thresholds as a function of age," *Albrecht Von Graefes Arch. Klin. Exp. Ophthalmol.*, vol. 215, no. 3, pp. 149–157, Jan. 1981.

[12] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, ITU-R Rec. BT 500-13, 2012.

[13] G. D. Boreman, *Modulation Transfer Function in Optical and Electro-Optical Systems*. Bellingham, WA: SPIE Press, 2001.

[14] N. Koren. (2011). *Understanding Image Sharpness Part 2: Resolution and MTF Curves in Scanners and Sharpening* [Online]. Available: http://www.normankoren.com/Tutorials/MTF2.html

[15] H. R. Taylor, "Applying new design principles to the construction of an illiterate E chart," *Am. J. Optom. Physiol. Opt.*, vol. 55, no. 5, pp. 348–351, May 1978.

[16] *Optics and Optical Instruments—Veiling Glare of Image Forming Systems—Definitions and Methods of Measurement*, ISO 9358:1994 Standard.

[17] *Photography—Electronic Still Picture Cameras—Resolution Measurements*, ISO 12233:2000.

[18] D. Williams and P. D. Burns, "Low-frequency MTF estimation for digital imaging devices using slanted edge analysis," in *Proc. SPIE/IS&T Symp. Electron. Imaging*, vol. 5294. 2004, pp. 93–101.

[19] K. Masaoka, M. Sugawara, and Y. Nojiri, "Multidirectional MTF measurement of digital image acquisition devices using a Siemens star," in *Proc. SPIE/IS&T Symp. Electron. Imaging*, vol. 7537. 2010, p. 75370 V.

[20] J. S. Lim, *Two-Dimensional Signal and Image Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1990.

[21] *Industrial Colour-Difference Evaluation*, CIE Publication 116-1995, 1995.

[22] P. G. J. Barten, *Contrast Sensitivity of the Human Eye and Its Effects on Image Quality*. Bellingham, WA: SPIE Press, 1999.

[23] M. Cowan, G. Kennel, T. Maier, and B. Walker, "Contrast sensitivity experiment to determine the bit depth for digital cinema," *SMPTE Motion Imaging J.*, vol. 113, no. 9, pp. 281–292, Sep. 2004.

[24] M. D. Fairchild and D. R. Wyble, "Mean observer metamerism and the selection of display primaries," in *Proc. 15th IS&T/SID Color Imaging Conf.*, Nov. 2007, pp. 151–156.

[25] J. J. Koenderink, A. J. van Doorn, and A. M. L. Kappers, "Pictorial relief," in *Seeing Spatial Form*, M. R. M. Jenkin and L. R. Harris, Eds. New York: Oxford Univ. Press, 2006, ch. 2, pp. 11–33.

[26] J. O. Drewery and R. A. Salmon, "Tests of visual acuity to determine the resolution required of a television transmission system," in *BBC R&D White Paper*, WHP 092. Sep. 2004.

[27] K. Masaoka, M. Emoto, M. Sugawara, and Y. Nojiri, "Contrast effect in evaluating the sense of presence for wide displays," *J. SID*, vol. 14, no. 9, pp. 785–791, Sep. 2006.

[28] L. L. Thurstone, "A law of comparative judgment," *Psychol. Rev.*, vol. 34, pp. 273–286, Jul. 1927.

[29] R. A. Bradley, "Paired comparisons: Some basic procedures and examples," in *Handbook of Statistics*, vol. 4, P. R. Krishnaiah and P. K. Sen, Eds. Amsterdam, The Netherlands: Elsevier, 1984, pp. 299–326.

[30] J. C. Handley, "Comparative analysis of Bradley–Terry and Thurstone–Mosteller paired comparison models for image quality assessment," in *Proc. IS&T PICS Conf.*, 2001, pp. 108–112.

[31] J. Morovič, "Psychovisual methods," in *Color Gamut Mapping*. Chichester, U.K.: Wiley, 2008, ch. 3, sec. 3.2, pp. 58–70.

[32] H. A. David, *The Method of Paired Comparisons*, 2nd ed. London, U.K./New York: Griffin/Oxford Univ. Press, 1988.

Kenichiro Masaoka (M'12) received the B.S. degree in electronics engineering and the M.S. degree in energy engineering from the Tokyo Institute of Technology, Tokyo, Japan, and the Ph.D. degree in engineering from the Tokyo Institute of Technology in 2009.

He joined Japan Broadcasting Corporation (NHK), Tokyo, in 1996. He is currently a Principal Research Engineer with the Advanced Television Systems Research Division, NHK Science and Technology Research Laboratories. He was with M. Fairchild and R. Berns for a six-month residency as a Visiting Scientist with the Munsell Color Science Laboratory, Rochester Institute of Technology, Rochester, NY, in 2012. His current research interests include color science, human vision, and digital imaging systems.

Dr. Masaoka is a member of the Institute of Image Information and Television Engineers of Japan.

Yukihiro Nishida graduated from Keio University, Tokyo, Japan, and received the M.Sc. degree in electrical engineering.

He is currently a Senior Research Engineer with the Advanced Television Systems Research Division, Japan Broadcasting Corporation Science and Technology Research Laboratories, Tokyo. He has been actively involved in the development of several broadcasting technologies, such as HDTV, digital broadcasting, video coding, multiplexing, and quality evaluation. One such area is the study of next-generation television systems such as UHDTV. He also plays an active role in the standardization of broadcasting technologies in the International Telecommunication Union Radiocommunication Sector (ITU-R), the Association of Radio Industries and Businesses, and the European Broadcasting Union. He is the Vice Chairman of the ITU-R Study Group 6 on Broadcasting Services and is also the Chairman of the Working Party 6B on Source Coding, Interfaces, Multimedia, and Interactivity. He was a Visiting Research Engineer with the Radiotelevisione Italiana Research Center, Turin, Italy, in 1992.

Masayuki Sugawara (M'03) received the B.S. and M.S. degrees in electric communication engineering, and the Ph.D. degree in electronic engineering from Tohoku University, Sendai, Japan.

He joined Japan Broadcasting Corporation (NHK), Tokyo, Japan, in 1983. Since 1987, he has been researching solid-state image sensor and HDTV camera with NHK Science and Technology Research Laboratories. He was an Associate Professor with the University of Electro-Communications, Tokyo, Japan, from 2000 to 2004. He is currently an Executive Research Engineer with NHK Science and Technology Research Laboratories and is engaged in the research of ultrahigh definition TV systems.

Dr. Sugawara is a member of SID, SMPTE, IEICE, and ITE. He has been attending ITU-R SG6 meeting since 2004 and has been active in the working parties that deal with program production.

Eisuke Nakasu (M'06) received the B.E. and M.E. degrees in electric engineering from Keio University, Tokyo, Japan, and the Ph.D. degree in engineering from the Tokyo Institute of Technology, Tokyo.

He joined Japan Broadcasting Corporation (NHK), Tokyo, in 1982. Since 1985, he has been in charge of the research and development of digital broadcasting systems, video coding, and picture quality evaluation. He is currently the Director of the Advanced Television Systems Research Division, NHK Science and Technology Research Laboratories.

Dr. Nakasu is a member of ITE and IEICE.

Yuji Nojiri received the Ph.D. degree in electrical engineering from Waseda University, Tokyo, Japan, in 2004.

He joined Japan Broadcasting Corporation (NHK), Tokyo, in 1978. Since 1981, he has studied HDTV and the various applications of this technology with NHK Science and Technology Research Laboratories (STRL). He developed the laser beam film recorder, telecine, and HDTV standards converter. From 1999 to 2001, he was with NHK Engineering Services, Inc., Tokyo, and joined the construction of the National TV Program archives. From 2001 to 2003, he studied on stereoscopic HDTV at the NHK STRL, and in 2004, he moved to the Division of Planning and Coordination. From 2005 to 2011, he studied high-resolution digital imageries. He is currently with NHK Integrated Technology, Inc., Tokyo. His current research interests include the application of high-resolution and large screen digital imageries.