Journal of Perceptual Imaging® 2(1): 010402-1–010402-11, 2019.

Relationship between Faithfulness and Preference of Stars in a Planetarium

Midori Tanaka▲
College of Liberal Arts and Sciences, Chiba University, Chiba, Japan
E-mail: midori@chiba-u.jp

Takahiko Horiuchi▲
Graduate School of Engineering, Chiba University, Chiba, Japan

Ken'ichi Otani
Engineering Division, Konica Minolta Planetarium Co., Ltd., Tokyo, Japan

Abstract. *A planetarium imitates a starry sky with physical and technical limitations using a dome, projector, and light source. It is widely used for entertainment, and astronomy and physics educations. In our previous study, we investigated the evaluation for faithful reproduction of a star field in a planetarium by performing psychometric experiments with 20 observers for plural projection patterns with different reproduction factors (color, luminance, and size of projected stars). In this study, we investigate the relationship between faithfulness and preference of a star field in a planetarium through a psychometric experiment with 47 observers. The experimental procedure followed the previous study. The rating of faithfulness improved for the projection pattern with a smaller star size. For the preference evaluation, the projection pattern with low luminance significantly lowered the preference rating. The results of the experiment indicate that the preferable star reproduction was different between male and female observers, whereas the faithful star reproduction was not significantly different in the evaluations between male and female observers. The male observers sought a faithful star reproduction as the preferred reproduction. In contrast, the female observers did not feel the faithful star reproduction preferable, and evaluated the more brilliant star reproduction as the preferred reproduction. These results were not dependent on the experience in astronomical observations.* © *2019 Society for Imaging Science and Technology.*
[DOI: 10.2352/J.Percept.Imaging.2019.2.1.010402]

1. INTRODUCTION

With the development of digital image devices such as displays and projectors, it has become possible to reproduce natural scenes on the devices. When the device can fully express the luminance and chromaticity of the natural scene, colorimetric color reproduction should be the optimal approach to represent the natural scene. However, luminance and chromaticity in the natural scene often exceed the limit of the device specification. For this purpose, various studies on image reproduction have been performed. For example, the reproductions of rich brightness and color in the natural scene have been studied through a high-dynamic-range (HDR) tone mapping method [1, 2] and gamut mapping method [3], respectively. For such image reproduction, we have to consider whether "faithful image reproduction" or "preferred image reproduction" is targeted, according to the application. In fields such as telemedicine, internet shopping, and digital archiving of arts, faithful image reproduction is required. In contrast, in fields such as skin image reproduction, preferred image reproduction is important [4]. Therefore, appropriate image reproduction methods have been developed according to the image application.

In the field of image application, there are fields where the appropriate image reproduction method has not been investigated, for example, the image reproduction of the star field in a planetarium. A planetarium imitates a starry sky with physical and technical limitations using a dome, projector, and light source; it is widely used for entertainment, astronomy and physics education, etc. In a planetarium, we can experience a similar perception for the real star observed from the earth from a downtown or place with small light pollution. We can also experience a preferred star-field reproduction with an excellent impression by projecting the innumerable stars, which could not be observed in reality. Although various image reproductions are available, the type of image reproduction desired for visitors in a planetarium has not been investigated in detail.

Star images in a planetarium are a set of spatially distributed point-like light sources with various colors, luminances, and sizes. The viewing environment is unique and different from other image reproduction situations, as a field of stars has to be projected onto a hemispheric screen representing the entire sky, which is different from general display devices such as plane screens with a limited field of view. In recent years, perceptual studies concerning image reproduction on new display devices such as curved displays [5–7] and head mounted displays [8–10] have been conducted, but to the best of our knowledge, similar studies using star images do not exist. Moreover, although a starry sky is three-dimensional, it is considered a two-dimensional

▲ IS&T Members.

Received Apr. 30, 2018; accepted for publication Jan. 23, 2019; published online Mar. 1, 2019. Associate Editor: Damon Chandler.
2575-8144/2019/2(1)/010402/11/$00.00

perception, approximately the same as a normal image, owing to the very large viewing distance between the stars and observers.

In addition, star images in a planetarium are mostly observed using scotopic vision, however, it is thought that the photopic or mesopic vision systems are also involved, as we can perceive star colors. Mesopic vision is active during the process of transition from photopic (illuminated) to scotopic (dark) vision, and utilizes a complicated mechanism for perceiving color in which cones and rods work together. For example, the peak of human spectral sensitivity to the perception of brightness shifts to shorter wavelengths with declining illumination [11]. Mesopic vision has special characteristics, such as the Purkinje shift, by which red colors (long wavelength) are perceived as darker than blue colors (short wavelength). In recent years, there have been studies concerning methods to reproduce images in the mesopic vision environment that consider optic nerve function [12–14]. There have also been psychophysical experiments which concern distinguishing colors in dim light (low illumination) environments [15, 16]. Thus, in a planetarium environment, where mesopic vision, scotopic vision, and rod intrusion interact in a complex matter, colorimetric reproduction is not necessarily the optimum reproduction. Owing to the above reasons, many of the findings from conventional experimental studies on the display of natural images may not be applicable to the reproduction of star images in a planetarium. Furthermore, the reproduction of star images in a planetarium has limited performances for color, luminance, contrast, resolution, and dome size, compared with general image output devices.

In our previous study [17], we investigated the evaluation of faithfulness of starry sky reproduction in a planetarium by performing psychometric experiments for plural projection patterns with different reproduction factors (color, luminance, and size of projected stars). We then analyzed the required factors for faithful image reproduction of a star field. However, we did not discuss whether the faithful image reproduction is preferable for observers. In this study, we investigate the relationship between faithfulness and preference of starry sky reproduction in a planetarium through a psychometric experiment.

2. EXPERIMENT

In order to investigate the factors that affect the evaluation of faithfulness and preference, we performed a psychometric experiment for projected star fields in a planetarium with different values of three parameters (color, luminance, and size) of individual stars. The basic experimental procedures follow our previous study, but some stimuli were changed from Ref. [17]. Twenty observers with astronomical observation experiences participated in the evaluation experiments. In this experiment, we targeted 47 adults with experience in astronomical observation, and who watch the actual starry sky on a daily basis.

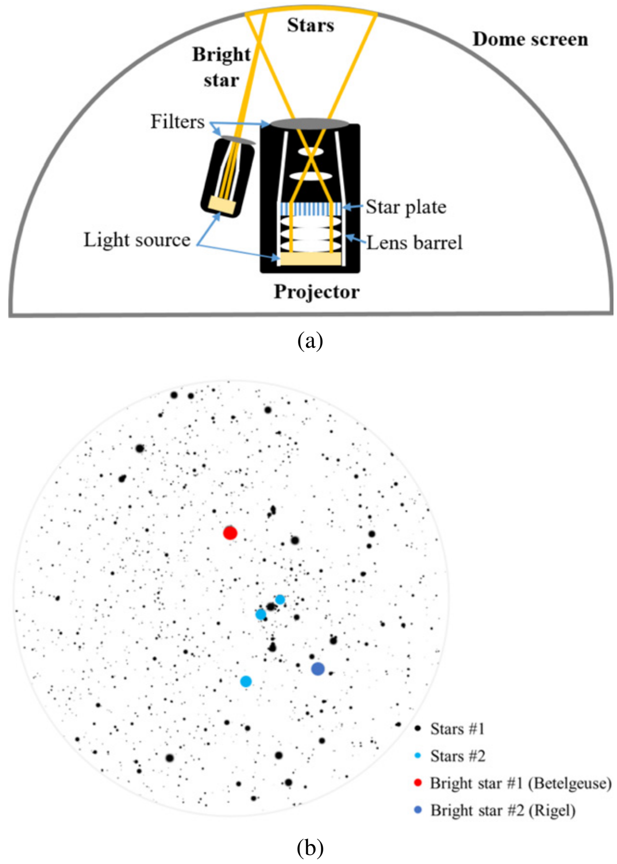

Figure 1. Star reproduction method. (a) System used to project star fields in a planetarium. (b) Representation of the color variation of stars used in our experiment.

2.1 *Experimental Setup*

We used the same reproduction method of stars and experimental stimuli as in the previous study [17]. Therefore, in this subsection, we briefly explain the reproduction method and star fields used as experimental stimuli.

2.1.1 *Reproduction Method of Stars*

We used an optical projection system in Figure 1(a) which generates a star image in a planetarium, as reproduced star images by a digital projection system are faced with more technical limitations related to providing a sufficient lightness, dynamic range, and minimum star size (depending on the resolution). In the optical projection system, fixed stars were projected on the dome screen by passing light through star plates installed in a lens barrel. The color and luminance of the stars were adjusted by inserting transmission filters in front of the light sources, while the star size was adjusted by varying the hole diameter of the star plates. For the representation of particularly bright stars, we developed another projection device to provide a sufficient dynamic range of luminance.

Table I. List of experimental projection patterns.

Projection Pattern	Changed Parameter	Remarks
Std	Standard	
C-1	Color-temperature shift	Pattern Std -100 mired
C-2	Color-temperature shift	Pattern Std +100 mired
C-3	Color-temperature shift	Pattern Std +200 mired
L-1	Luminance shift	1/2× Pattern Std
L-2	Luminance shift	2× Pattern Std
S-1	Size	2/3 of Pattern Std
S-2	Size	3/2 of Pattern Std
B	Color	Planckian locus

2.1.2 *Experimental Stimuli*
Stars around Orion (1/32 area of the entire sky) were selected as experimental stimuli to evaluate the faithfulness and preference of the star field in the planetarium. This area originally includes Sirius, Milky Way, and some nebulas; however, we excluded them as projection targets, as Sirius is the brightest star and may produce a bias during the evaluations owing to its attractiveness. The Milky Way and nebulas were excluded owing to the difficulty of controlling the hole diameter to represent them. As a preliminary experiment, we conducted an evaluation using the Std pattern, including the Milky Way and Nebula, and found that they did not affect the evaluation results. Therefore, in our opinion, their inclusion in other patterns would not have a significant effect on the results. In addition, we excluded external factors such as the atmospheric extinction affecting the twinkling of stars and airlight from the solar system (zodiacal light), and airlight from the Earth's atmosphere (light pollution) to provide a stable star image, according to the previous study. Fig. 1(b) shows the projected star image used as the experimental stimulus in our experiment. The colors of the stars around Orion were represented using four different color reproduction methods, which consisted of two types of fixed star color (Stars #1 and #2) and two types of bright star color such as those of Betelgeuse and Rigel (Bright star #1 and #2). This color reproduction method was designed according to the magnitude and color temperature of each star. As the experimental stimulus, we used a set of 1,378 fixed stars and two additional bright stars with magnitudes below 7.4 considering the visibility limitation [18].

We used 9 types of experimental patterns obtained by changing the color, luminance, and projection size with respect to the whole star image. In our previous study, we prepared three patterns by changing the luminance; however, the evaluation value of the brightest pattern (L-3 in Ref. [17]) was remarkably low. Therefore, the pattern L-3 was excluded from this experiment. Neutral-density (ND) and color-temperature-change filters were used to change the star's color and luminance. The list of projection patterns used as experimental stimuli is shown in Table I. Figure 2(a) shows the measured chromaticity values of the projected stars, while Fig. 2(b) shows the luminance of the stars in the standard pattern and L-pattern shifts in luminance. The colors of these stars projected onto the dome screen were measured using a spectroradiometer (CS-2000, Konica Minolta). The main points of the reproduction method for each pattern are:

- Std: A standard pattern was designed to provide an equivalent perception of major stars obtained from an actual starry sky by experienced observers with sufficient experience in astronomical observations.
- Pattern C: Three C-patterns (C-1, C-2, and C-3) had color-temperature shifts of -100,$+100$, and $+200$ mired, respectively, from the standard pattern obtained using color conversion filters. In addition to the patterns C-1 and C-2 (± 100 mired), the pattern C-3 (a shift of $+200$ mired with a strong reddish color) was provided as the visual sensitivity at long wavelengths becomes weak owing to the Purkinje effect in mesopic vision.
- Pattern L: We changed the contrast with the background of the night sky and visibility by preparing L-patterns. The luminances of the stars in the patterns L-1 and L-2 were approximately 0.5 and 1.9 times larger than those of the standard pattern, respectively. The luminance in these L-patterns was changed using ND filters; however, they retained the color and size of the standard pattern.
- Pattern S: We controlled the projection size of the stars by expanding or contracting the hole diameter of the star plate for the standard pattern. The luminance and color of the stars in the S-patterns were the same as those of the standard pattern, but their brightness, as perceived by observers, was made different from that of the standard pattern by changing the projection size. Compared with the standard pattern, the patterns S-1 and S-2 had disc area ratios of 2/3 and 3/2, respectively.
- Pattern B: A Planckian locus pattern with a physically approximated chromaticity of the Planckian locus was prepared, as the actual star color had the chromaticity of its locus. We used the color index of a star catalog to design the star color of the experimental stimulus [19, 20]. While pattern C was prepared by inserting the color-temperature conversion filters onto the Std pattern for shifting the color temperature of four stars together, the star color in pattern B was separately controlled to reproduce the desired chromaticity along Planckian locus which was independent to Std pattern.

2.2 *Evaluation Method*
We performed a psychometric experiment to evaluate the perception of faithfulness and preference of star-image reproduction in a planetarium for 9 types of projection patterns, as summarized in Table I. In the faithfulness evaluation, the observers evaluated the faithfulness using an opposite word pair ("faithful" / "non-faithful") and five integer levels from -2 to $+2$ compared with the actual starry sky in their memory, and wrote their evaluation

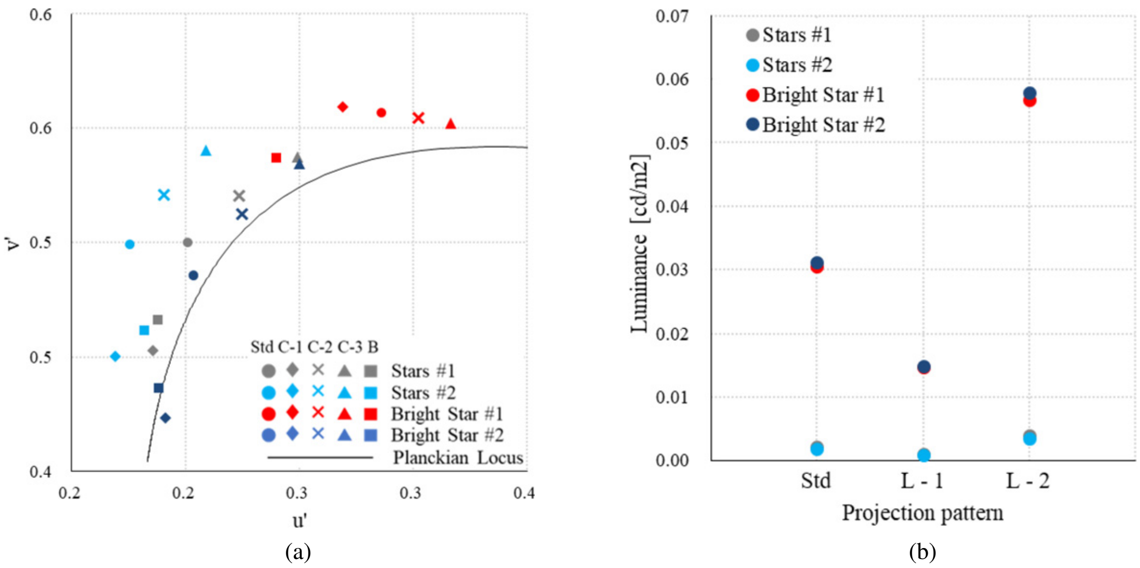

Figure 2. Conditions of the star reproduction. (a) Star colors of Std, C-, and B-patterns. (b) Star luminances of Std and L-patterns.

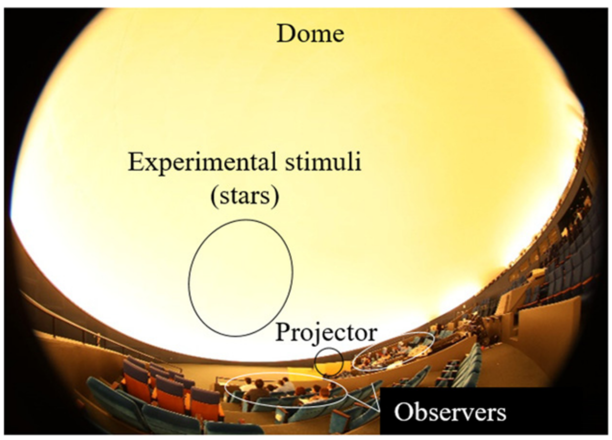

Figure 3. Snapshot of the experimental environment (captured at a bright dome).

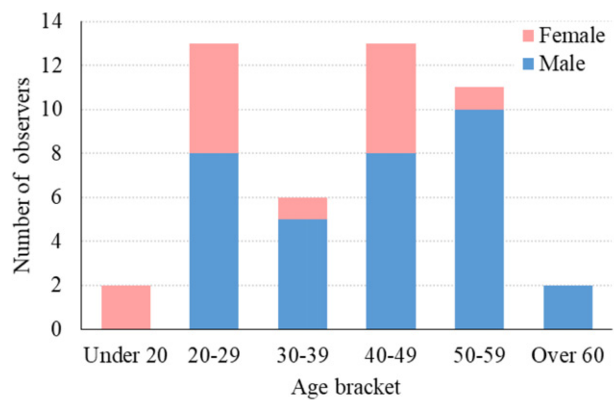

Figure 4. Numbers of observers according to their age.

values down on answer cards based on a 5-point Likert scale. The meanings of the evaluation levels were: −2 (not faithful), −1 (slightly not faithful), 0 (neutral), +1 (slightly faithful), and +2 (faithful). In the preference evaluation, the observers evaluated using another opposite word pair ("preference" / "non-preference") with the same five integer levels from −2 to +2 without a comparison target. The answer task was performed in complete darkness, only with the projected star image to maintain the dark adaptation; there was no other bias to discriminate against particular answers. In the evaluation, there was no designated fixation point; the observers were able to freely observe the star image. Therefore, they could well judge the color and brightness of the whole projection stimuli using the foveal vision by the cones and peripheral vision by the rods. A snapshot image of the illuminated dome captured using a fish-eye lens is shown in Figure 3.

This psychometric experiment was performed using the dome of the planetarium. Each star pattern was projected to the position of the oval mark in the figure. The diameter of the dome screen was 23 m; the zenith of the dome screen was slanted 15° to the front. There was no other illumination in the space where the experiments were performed, aside from that of the projected starry sky image. We confirmed that the room appeared completely dark. It was not possible to verify the low light level using the CS-2000 as it was too dark to measure (< 0.003 cd/m2). The average viewing distance between the observers and center of the projected star-field image was approximately 10.7 m (viewing angle: 37.3°); it slightly varied according to the seat position.

A total of 47 observers participated in this experiment. Biased observers participated in our previous study in terms of gender and age; in contrast, in this experiment, various

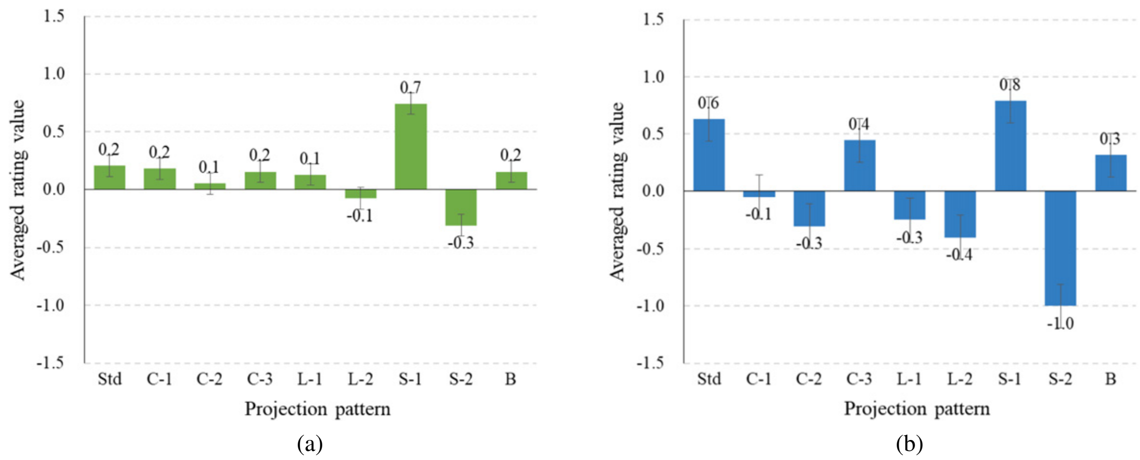

Figure 5. Faithfulness evaluation results for each projection pattern. (a) Present experiment. (b) Previous experiment.

types of observers participated. As shown in Figure 4, 33 male and 14 female observers with different ages participated. A requirement for the observers was that they had experience in astronomical observation. Two experiments were conducted, for a set of 26 observers, and a set of 21 observers, respectively, in order to maintain a similar viewing distance between the observers and the projection on the planetarium dome. The experimental process was the same for both sets of observers. Once the observers took a seat in the planetarium, the illumination of the dome was turned off. We confirmed that the brightness of the dark dome was lower than a magnitude of 23.0 using the sky quality meter. Most stars with brightness higher than a magnitude of 7.4 reproduced by the projection could be perceived as the magnitude limit for observation in the 23.0-magnitude darkness was approximately 7.0 [18].

Prior to the psychometric experiment, the observer received instructions for the evaluation method and exercised using a practice pattern for 20 min. It was assumed that the observers had completed the dark adaptation by this time. The five-point evaluation system for rating the faithfulness (or preference) of the projection on a scale of -2 to $+2$ was explained to the observers. The observers were provided with an answer card to record their ratings. During the experiments, the observers were not allowed to talk to each other or to move from their appointed seat. The experiment was proceeded by oral instructions using a microphone in the dark dome. In this experiment, 10 randomly projected patterns (including standard patterns twice to confirm reproducibility) were used. Note that the first Std pattern was projected first. The observers evaluated the faithfulness and preference of each pattern (each of them within 15 s) after observing the star image for 30 s. For the observation, we set a time period of 30 s as a previous study reported that the detection limit of the brightness in the dark environment became constant after 15 s during the experiment [21]. A short break of few minutes was set between the pattern evaluations to reset the last evaluation on the observers' perception and exchange the projection pattern. The observers were given a ten-minute break after half of the experimental patterns had been evaluated, and the experiment was resumed after a period of dark adaptation. All experiments were performed in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki). Written informed consents were obtained from all of the participants.

3. EXPERIMENTAL RESULTS

The significance of the answered evaluations for each pattern was confirmed using a t-test after excluding the outlier data using the Smirnoff–Grubbs test and verifying the distribution equality of evaluations by an F-test. Here, we assumed the normality of the evaluated data. These tests were conducted separately for all 36 pattern combinations ($9 \times 8/2$) for each evaluation index (faithfulness and preference). Seven observers (six males, one female) and six observers (three males, three females) of 47 total observers were excluded from the evaluations of faithfulness and preference data, respectively. There was no significant difference in the answer results for the standard pattern, which was evaluated twice to confirm the reproducibility. Therefore, we used the second scores of the standard pattern for the analysis. The intra-observer variances (calculated from the evaluation of the two passes with the standard patterns) and inter-observer variances (calculated from the evaluations of all of the projected patterns) were 0.35 and 1.16 for the faithfulness evaluation, and 0.40 and 1.25 for the preference evaluation, respectively.

3.1 *Faithfulness Evaluation*

First, we checked whether the faithfulness evaluation for each projection pattern satisfied the results in Ref. [17], as we used observers with different characteristics from those of the previous experiments. Figure 5(a) shows the average rating

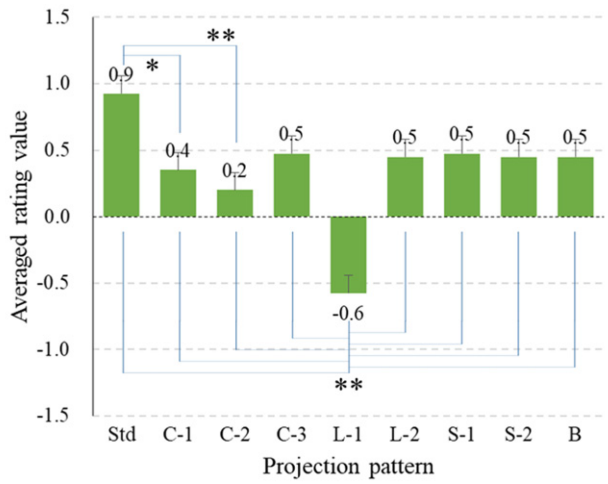

Figure 6. Preference evaluation results of all of the observers for each projection pattern.

value (with the standard error among observers) for each pattern. For comparison, we show the results of Ref. [17] in Fig. 5(b). As shown in the figure, except for the projection pattern S-2 ($p < 0.05$), there was no significant difference in each projection pattern between the previous and present data. What is different from the subjects in the previous study was that the subjects in the present study were watching the actual starry sky on a daily basis as described in Section 2. Therefore, the rating became sensitive to the faithfulness of brighter stars, and the evaluation for the pattern S-2 had might be significantly worse. Even for S-2, the average rating values of the previous and present experiments were −1.0 and −0.3, respectively; the tendencies of negative evaluation were consistent. These results confirmed that the rating values for faithful reproduction in this experiment supported the results in Ref. [17].

The projection pattern with the highest rating of faithfulness was the pattern S-1, as shown in Fig. 5. There was a significant difference between the pattern S-1 and all other patterns. In addition, the significant difference between the Std and S-2 patterns with a negative rating value was confirmed. On the other hand, the C- and L-patterns did not have a significant difference with another pattern. These results indicate that the projection size significantly influenced the faithfulness evaluation. In contrast, the tolerance to lack of faithfulness was high regarding the color-temperature shift in the prepared patterns used in this experiment, as these patterns had no different rating values.

We can summarize the conclusions on faithful reproduction as follows. In addition, in the faithfulness evaluation for the projection pattern with a smaller star size, the rating of faithfulness improved. In contrast, the rating was remarkably negative when the pattern had stars with a larger size than that in the standard pattern. Regarding the shifts in the color temperature of the stars, the observers could distinguish differences in color but did not give negative ratings for changes in the color pattern.

3.2 *Preference Evaluation*

In this subsection, we consider the preferred reproduction. The preference evaluation was also performed by 47 observers. Figure 6 shows the average rating value with the standard error among the 41 observers excluded outliers for each pattern. A significant difference is indicated by symbols (*: $p < 0.05$, **: $p < 0.01$). The projection pattern with the highest rating of preference was the pattern Std, as shown in Fig. 6. There was a significant difference between the patterns C-1 and C-2. In contrast, the pattern L-1 had the lowest rating of preference with a larger significant difference with all of the other patterns ($p < 0.01$). On the contrary, color or size change patterns or brighter patterns such as the C-3, B, S-1, S-2, and L-2 patterns had similar ratings without a significant difference with the highest-rated pattern Std. These results indicate that various observers accepted the Std pattern

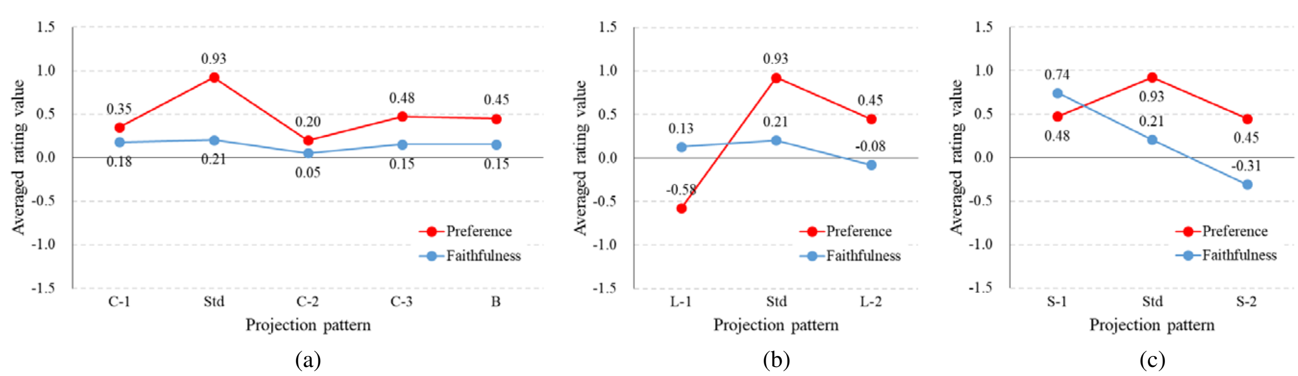

Figure 7. Total average rating values of preference and faithfulness evaluations. (a) Std and pattern C. (b) Std and pattern L. (c) Std and pattern S.

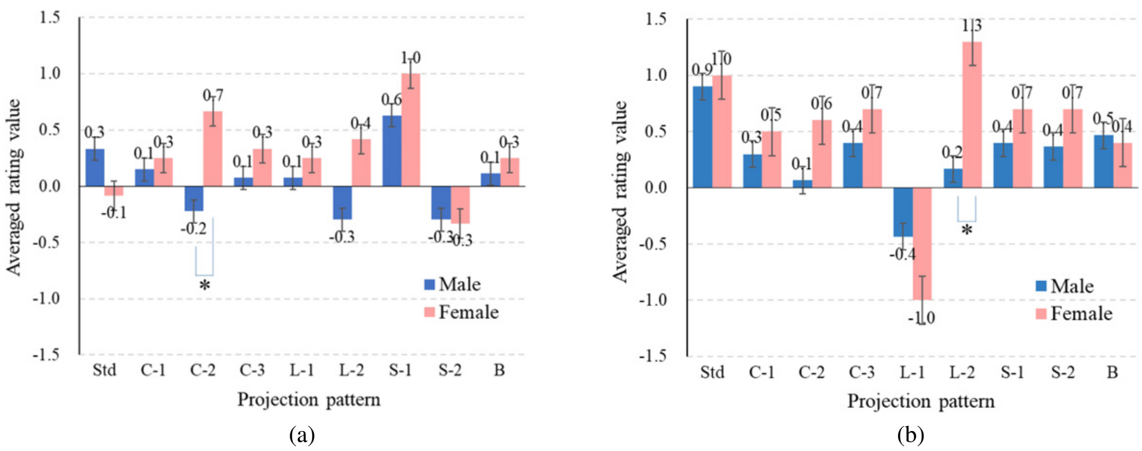

Figure 8. Gender differences of average rating values for preference and faithfulness evaluations. (a) Faithfulness. (b) Preference.

as the preference star reproduction, which was prepared considering the perceptual equivalence to the appearance of the actual starry sky. On the contrary, the darkness in the star reproduction sensitively negatively influenced the preference evaluation.

3.3 *Relationship of the Evaluations between Faithfulness and Preference*

In this subsection, we investigate the relationship between faithfulness and preference of stars in a planetarium.

3.3.1 *Comparison of Evaluations for Faithfulness and Preference*

Figure 7 summarizes the average rating values for faithful and preferred reproductions of each projection pattern. Patterns L-1, L-2, and S-2 were significantly different ($p < 0.05$) between the two evaluations. Regarding the luminance variation pattern, the rating value of the preferred reproduction fluctuated, compared with the evaluation of the faithful reproduction. It is worth noting that compared with the rating values of faithful reproduction, the rating values of preferred reproduction decreased when the stars were dark and increased when the stars were bright.

Regarding the size variation pattern, with the increase of the size, the rating value of faithful reproduction monotonously decreased; however, the rating value of preferred reproduction was almost the same between the patterns S-1 and S-2; i.e., although the large-size-star image was not faithful to the actual star field, it was evaluated as preferable. For the evaluation of the color pattern C, there was no significant difference of rating values between the faithful and preferred reproductions.

The faithfulness score for Std was similar to the average for all the patterns, but it had the second highest score overall, behind the S-1 pattern. This is because the reproduction S-1, with a size equivalent to 2/3 the size of Std, resembled the star size in the genuine starry sky. However, as the number of visible stars in the S-1 pattern decreased compared with the Std pattern, the preference score of S-1 decreased, and the Std pattern received the highest preference score.

3.4 *Gender Evaluation Differences*

The star-image analysis in Section 3.3.1 revealed differences between the faithfulness and preference evaluations in the three patterns (L-1, L-2, and S-2) for luminance and size. We further revealed that there was a characteristic gender difference.

The gender differences of the average rating values for preference and faithfulness evaluations are shown in Figure 8. The error bars represent the standard errors among each gender (male: 33 observers, female: 14 observers) for each pattern. As shown in Fig. 8(a), the evaluation of faithfulness seems to vary between male and female observers; however, there was not a statistically significant difference, except for the projection pattern C-2 ($p < 0.05$). The evaluation of the pattern C-2 seems to be a convincing result, as, according to Ref. [17], the evaluation was divided into a positive and negative evaluation group. Regarding the preference evaluation, as shown in Fig. 8(b), the tendency of evaluation, positive (plus rating) and negative (minus rating), was consistent between male and female observers, and there was no statistically significant difference except for the evaluation of the pattern L-2 ($p < 0.05$). It is worth noting that for the female observers, the projection patterns with high brightness such as the patterns L-2 and S-2 were highly rated as preferred.

Further, within each category of male and female observers, we investigate the relationship between faithfulness and preference evaluations. Figure 9 shows the differences in rating values between the two evaluations. As shown in Fig. 9(a), for the male observers, there was no significant difference in any pattern. This suggests that the male observers expected planetariums with faithful reproduction as a preferred reproduction. In contrast, as shown in

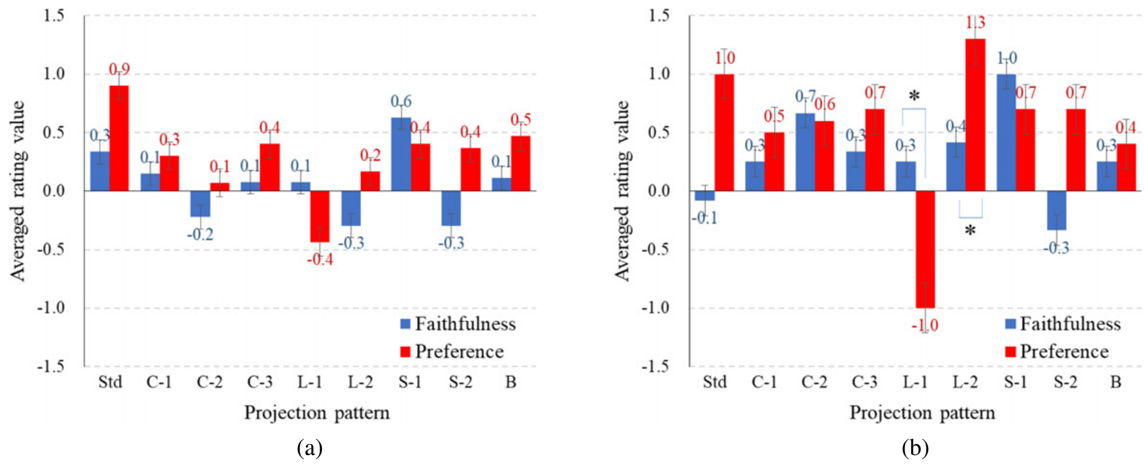

Figure 9. Relationships between faithfulness and preference evaluations for each gender. (a) Male. (b) Female.

Fig. 9(b), the female observers had different impressions. The evaluations by the female observers were significantly different for two patterns (L-1 and L-2; $p < 0.05$). They preferred the high-luminance reproduction rather than the faithful reproduction. In other words, they expected the planetarium to reproduce the brilliant star.

Incidentally, we investigated the experience of astronomical observation by conducting a questionnaire after the experiments. We analyzed experience using observers' data about years of astronomical observation experience, the frequency of observation, and locations where the sky was observed, etc., using several analysis methods including correlation and clustering. However, we did not find a relationship between observers' ratings and their experience. Therefore, we concluded that the results of the gender differences did not depend on the experience in astronomical observations.

Figure 10 shows the rating distributions of faithfulness and preference evaluations for each projection pattern, for the male (Fig. 10(a)) and female (Fig. 10(b)) observers. (As a Figure A1 in Appendix, we show the gender difference of rating distributions between faithfulness and preference evaluations for each pattern.) The size of the circle represents the answer ratio for each pattern. From the upper-left panel to the bottom-right panel, the total distribution for all patterns, Std and color variation patterns, Std and luminance variation patterns, and Std and projection size variation patterns are shown, respectively. In addition, a coefficient of correlation calculated by Pearson correlation with/without the significance between the evaluations for faithfulness and preference is shown in each figure. The total distribution had a significant correlation (**: $p < 0.01$) for both genders; the correlation coefficient for the male observers ($N = 33$) was higher than that for the female observers ($N = 14$), despite the more than twice larger number of male observers than that of female observers. Furthermore, the results of the male observers had significant correlations for each variation pattern ($p < 0.01$). In contrast, a significant correlation was confirmed only for the size variation patterns in the results of the female observers ($p < 0.05$). In particular, the faithful and preferred reproductions were not correlated for the luminance variation pattern. These results support the results in Fig. 9; i.e., the faithful and preferred reproductions were consistent for the male observers, while for the female observers, they differed, in particular, with respect to the luminance evaluation. Figures 11(a) and (b) show the inter-participant correlation per pattern sorted in ascending order w.r.t. the correlation. As shown in the figure, in the case of a male observer, a significant correlation is confirmed between the evaluation of faithfulness and preference in all patterns. In contrast, in most patterns, there is no significant correlation, especially correlation with pattern L cannot be confirmed. These results also support the results that the faithful and preferred reproductions were consistent for the male observers, while for the female observers, they differed with respect to the luminance evaluation.

4. CONCLUSIONS

We performed psychometric experiments to investigate the factors required to reproduce faithful and preference star-field images in a planetarium. We projected stars as experimental stimuli with different characteristics for three types of parameters (color, luminance, and size). A total of 47 observers participated in the evaluation experiments for 9 types of star-image patterns projected on a dome screen in a planetarium.

We checked whether the faithfulness evaluation for each projection pattern satisfied the results in a previous study, as we used observers with different characteristics from those of the previous experiments. The rating of faithfulness improved for the projection pattern with a smaller star size. The results confirmed that the rating values for faithful reproduction in this experiment supported the

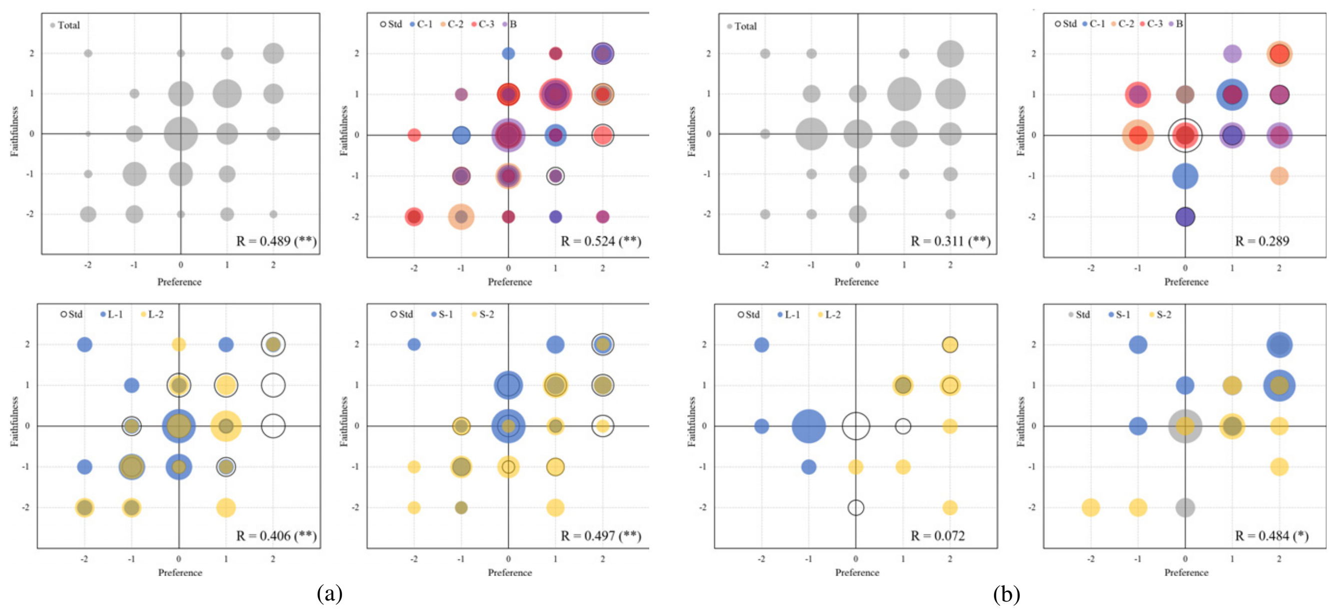

Figure 10. Rating distributions of faithfulness and preference evaluations for total rating, color-changed patterns and Std, luminance-changed patterns and Std, and size-changed patterns and Std. (a) Results evaluated by the male observers. (b) Results evaluated by the female observers.

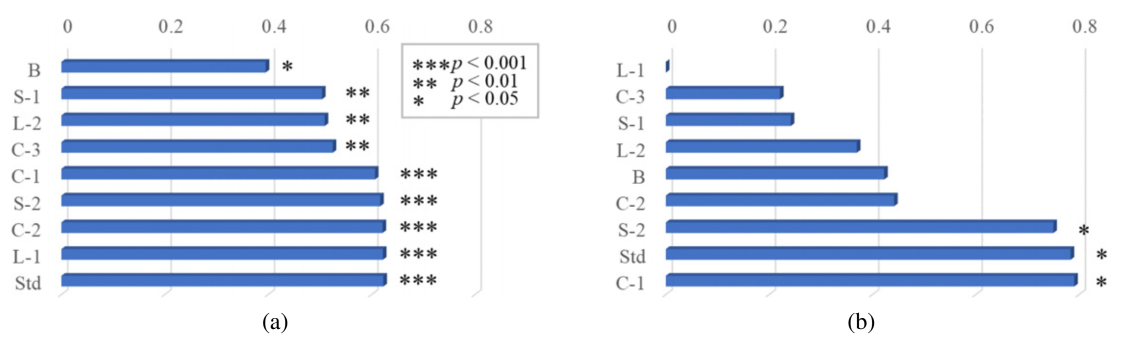

Figure 11. Inter-participant correlation per pattern. (a) Male. (b) Female.

results in the previous study. For the preference evaluation, the projection pattern with low luminance significantly lowered the preference rating. We then analyzed the relationship between the faithful and preferable evaluations. For the male observers, there was no significant difference in any projection pattern. This suggested that the male observers expected planetariums with faithful reproduction as a preferable reproduction. In contrast, for the female observers, the evaluation was significantly different in the luminance variation patterns. They preferred a brighter reproduction rather than a faithful reproduction. In other words, they expected the planetarium to reproduce the brilliant star. These gender-difference results did not depend on the experience in astronomical observations.

Following our experimental results, we interviewed a planetarium management company in Japan. The current projections in the planetarium have one of two purposes. The first purpose is for entertainment, which requires a pleasant reproduction of the night sky, and the second purpose is astronomical education which requires a faithful reproduction of the real starry sky. Through our investigation, the factors required to create a pleasant projection (preference), and one faithful to a genuine star field (faithfulness), in a planetarium setting were clarified.

APPENDIX.

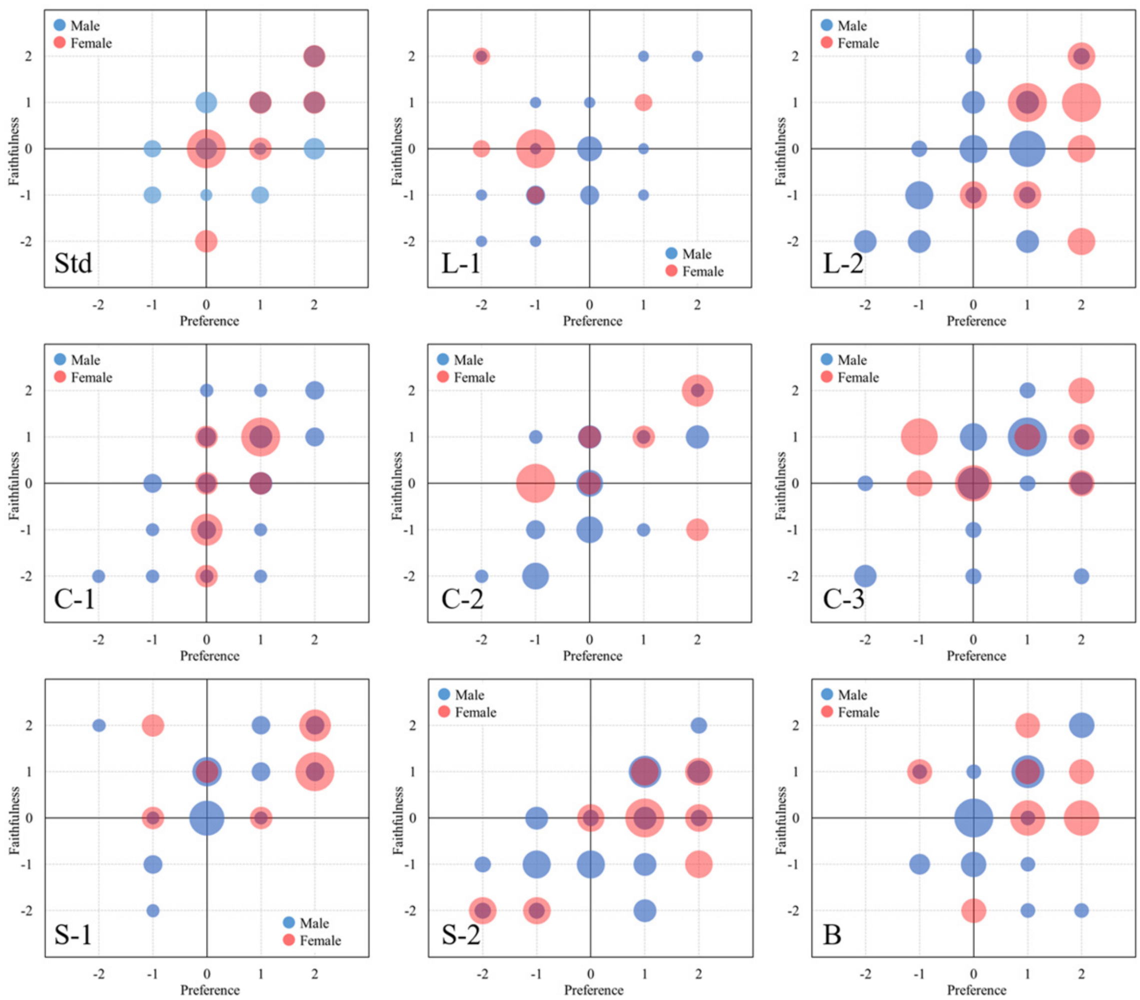

Figure A1. The gender difference in rating distributions between faithfulness and preference evaluations for each pattern.

REFERENCES

1. G. Eilertsen, R. K. Mantiuk, and J. Unger, "A comparative review of tone-mapping algorithms for high dynamic range video," Computer Graphics Forum **36**, 565–592 (2017).
2. X. Cerda-Company, C. A. Parraga, and X. Otazu, "Which tone-mapping operator is the best? A comparative study of perceptual quality," J. Opt. Soc. Am. A **35**, 626–638 (2018).
3. J. Morovic and R. M. Luo, "The fundamentals of gamut mapping: A survey," J. Imaging Sci. Technol. **45**, 283–290 (2001).
4. R. W. G. Hunt, I. T. Pitt, and L. M. Winter, "The preferred reproduction of blue sky, green grass and caucasian skin in colour photography," J. Photogr. Sci. **22**, 144–150 (2016).
5. T. Burney and P. Lock, "Measuring game-play performance and perceived immersion in a domed planetarium projection environment," *Springer Series in Lecture Notes in Computer Science* (Springer, Berlin, Germany, 2007), Vol. 4740, pp. 22–27.
6. N. Na, K. A. Jeong, and H.- J. Suk, "Do curved displays make for a more pleasant experience?," *IS&T Electronic Imaging: Human Vision and Electronic Imaging XX Proc.* (IS&T, Springfield, VA, 2015), Vol. 9394, 939419.
7. S. Mun, M.-C. Park, and S. Yano, "Evaluation of viewing experiences induced by a curved three-dimensional display," Opt. Eng. **54**, 103104 (2015).
8. J. P. Rolland, W. Gibson, and D. Ariely, "Towards quantifying depth and size perception in virtual environments," Presence: Teleoperators Virtual Environ. **4**, 24–49 (1995).
9. H. Ha, Y. Kwak, H. Kim, Y.-J. Seo, and W.-S. Park, "The preferred luminance of head mounted display (HMD) over time under two different surround conditions," *Proc. IS&T CIC25: Twenty-Fifth Color and Imaging Conf.* (IS&T, Springfield, VA, 2017), pp. 286–289.
10. P. J. Pardo, M. I. Suero, and Á. L. Pérez, "Correlation between perception of color, shadows, and surface textures and the realism of a scene in virtual reality," J. Opt. Soc. Am. A **35**, B130–B135 (2018).
11. J. A. S. Kinney, "Comparison of scotopic, mesopic, and photopic spectral sensitivity curves," J. Opt. Soc. Am. **48**, 185–190 (1958).
12. A. G. Kirk and J. F. O'Brien, "Perceptually based tone mapping for low-light conditions," *ACM SIGGRAPH* (ACM, New York, NY, 2011), pp. 1–10.
13. M. Mikamo, B. Raytchev, T. Tamaki, and K. Kaneda, "A tone reproduction operator for all luminance ranges considering human color perception," *Eurographics* (The Eurographics Association and John Wiley & Sons, Ltd., Hobokeny, NJ, USA, 2014).
14. J. C. Shin, N. Matsuki, H. Yaguchi, and S. Shioiri, "A color appearance model applicable in mesopic vision," Opt. Rev. **11**, 272–278 (2004).

[15] J. Pokorny, M. Lutze, D. Cao, and A. J. Zele, "The color of night: Surface color perception under dim illuminations," Vis. Neurosci. **23**, 525–530 (2006).
[16] J. C. Shin, H. Yaguchi, and S. Shioiri, "Change of color appearance in photopic, mesopic and scotopic vision," Opt. Rev. **11**, 265–271 (2004).
[17] M. Tanaka, T. Horiuchi, K. Otani, and P. Hung, "Evaluation for faithful reproduction of star fields in a planetarium," J. Imaging Sci. Technol. **61**, 060401 (2017).
[18] B. E. Schaefer, "Telescopic limiting magnitudes," Astron. Soc. Pac. **102**, 212–229 (1990).
[19] M. A. C. Perryman, L. Lindergren, L. Kovalevsky, E. Hoeg, U. Bastian, P. L. Bernacca, M. Crézé, F. Donati, M. Grenon, M. Grewing, F. can Leeuwen, H. van der Marel, F. Mignard, C. A. Murray, R. S. Le Poole, H. Schrijver, C. Turon, F. Arenou, M. Froeschlé, and C. S. Petersen, "The hipparcos catalogue," Astron. Astrophys. **323**, L49–52 (1997).
[20] G. H. Jacoby, D. A. Hunter, and C. A. Christian, "A library of stellar spectra," Astrophys. J. Suppl. Ser. **56**, 257–281 (1984).
[21] H. R. Blackwell, "Contrast Thresholds of the Human Eye," J. Opt. Soc. Am. **36**, 624–643 (1946).