https://doi.org/10.2352/CIC.2025.33.1.6

Testing the Acceptability Prediction of Color Difference Models using the Multiple Illumination Scenarios (MIS) Dataset

*Zhiyu Chen[1], Chenyu Wang[1], Qiang Xu[2], Qiang Liu[1] **
[1]Academy of Advanced Interdisciplinary Studies, Wuhan University, Wuhan, China
[2]Honor Device Co., Ltd., Shenzhen, China
**Corresponding author: Qiang Liu, liuqiang@whu.edu.cn*

Abstract

With the rapid advancement of mobile imaging and the increasing demand for perceptually accurate white balance (WB) algorithms, the need for a comprehensive dataset providing perceptual acceptability assessments across diverse illumination conditions has arisen. To address this gap, we constructed the Multiple Illumination Scenarios (MIS) dataset, which spans both pure colors and complex objects under single and multiple illuminant conditions. Observer-based acceptability ratings were collected and analyzed across 3,465 trials, revealing heightened sensitivity to chromatic deviations in regions of low lightness and chroma. Additionally, spatial and illuminance factors were found to modulate color acceptability judgments in multi-illuminant scenarios. Based on these findings, we proposed two new metrics to improve the performance of current color difference models: one weighted by color appearance attributes and another that incorporates spatial and illuminance factors. Evaluation results demonstrated that our proposed metrics showed improved correlation with perceptual judgments across all tested color difference models. By incorporating more realistic datasets and integrating alternative WB error evaluation metrics, we aim to advance research into the prediction of WB error acceptability under complex lighting environments.

Introduction

With the rapid advancement of mobile imaging technology and the growing demand for high-fidelity visual reproduction, white balance (WB) algorithms have received increasing attention across a wide range of imaging applications and illumination scenarios [1–6]. These algorithms aim to estimate the color of the illuminant present in a scene and correct chromatic casts accordingly [6–8]. Some classical methods demonstrate robust performance under single-illuminant conditions, while their accuracy often degrades in complex environments such as multi-illuminant scene [4,9]. Thus, to effectively correct such multiple illuminations, several methods have been proposed and evaluated [5,6,10].

Current evaluation protocols for WB algorithms predominantly measure chromaticity differences between the estimated and ground-truth illuminants [11–13]. Commonly adopted metrics include recovery angular error [14], reproduction angular error [15], and Euclidean distance [16]. However, these approaches typically consider only chromaticity deviations, neglecting luminance variations caused by different illuminance levels. Moreover, they are often insufficient for accurately evaluating performance in aforementioned challenging conditions like nonuniform or multiple illuminant scenarios.

An alternative and perceptually more intuitive method to assess image reproduction quality is through inter-image color differences. This approach allows for the incorporation of both chromatic and luminance cues, making it more applicable to heterogeneous lighting contexts. Central to this approach is the selection of an appropriate uniform color space or color difference formula. Although various color difference formulas have been proposed [17–20], most are derived from experiments with simple color patches under controlled viewing conditions [21]. Consequently, their predictive capability in complex, real-world scenes remain unexamined.

Moreover, prior work has largely concentrated on color perceptibility, i.e., the threshold at which differences become noticeable or discriminable [17,20,22], rather than color acceptability, which reflects subjective tolerance to chromatic deviations that is particularly relevant to industrial applications [23]. While studies have demonstrated that perceptibility and acceptability are not equivalent [24], to the best of our knowledge, no comprehensive dataset exists that provides perceptual acceptability assessments across diverse illumination conditions.

To address this gap, we presented the Multiple Illumination Scenarios (MIS) dataset, comprising observer-based acceptability ratings for scenes with a wide range of illumination configurations, including single-illuminant scenarios with pure colors and complex objects, colored-light scenarios, and multiple illuminant scenarios. Based on this dataset, we systematically evaluate the predictive performance of established color difference models. Additionally, optimized algorithms were also proposed for general and multi-illuminant conditions to improve the robustness of perceptual acceptability predictions in practical WB applications.

Methods

Multiple Illumination Scenarios (MIS) Dataset

The proposed Multiple Illumination Scenarios (MIS) dataset comprises three representative scene categories that reflect typical imaging conditions encountered in real-world applications. Firstly, scenarios with pure colors and single illuminant (PSI) are representing contexts like painted surfaces with uniform color distributions, which was considered as one of the crucial aspects of WB algorithm [25,26]. Secondly, scenarios with complex objects and single illuminant (CSI), representing the typical contexts including complex color distributions that has been extensively researched. And finally, the multiple illuminants (MI) scenarios simulated real-world conditions like mixed indoor lighting, including combinations of different areas, illuminant colors, illuminance levels together with high dynamic range (HDR) scenes.

For pure colors and single illuminant (PSI) scenarios, fifty-two color centers were selected as the reference stimuli, covering neutral, low- and medium-saturation, and several high-saturation colors, as in Fig. 1(a). For each center, 32 systematically deviated testing stimuli were generated, excluding those beyond the display gamut. Fig. 1(b) exhibits the example of the neutral reference color center together with the testing stimuli.

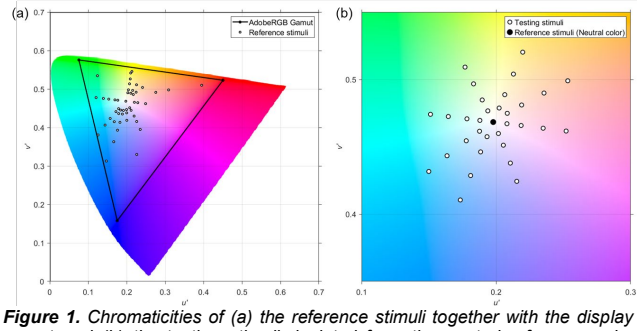

Figure 1. *Chromaticities of (a) the reference stimuli together with the display gamut and (b) the testing stimuli deviated from the neutral reference color plotted on the CIE 1976 u'v' chromaticity diagram.*

The complex objects and single illuminant (CSI) scenarios utilized ColorChecker chart to represent diverse colors in general contexts while maintaining controlled condition. The overall procedure of constructing this dataset were shown in Figure 2.

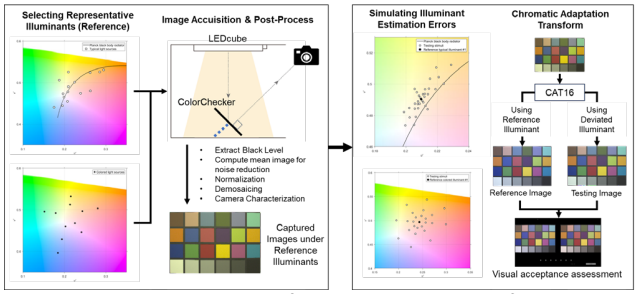

Figure 2. *The general procedure of constructing the dataset of scenarios with complex objects and single illuminant (CSI).*

Fifteen canonical illuminants were selected, including representative natural and artificial light sources (including domestic contexts, shops, and workplaces) sourced from existing literature [27–29], spanning a wide range of correlated color temperatures (CCTs). Additionally, owing to the spectral tunability of LED light sources, the chromaticity of artificial indoor lighting often deviated from the Planckian and Daylight locus and exhibited negative Duv values (Distance deviated from the Planckian locus) to achieve better color rendition quality [28].These illuminants were physically recreated using a programmable LEDCube light booth, where the spectrum for each target chromaticity was determined by optimizing the LED channels to match CIE u'v' coordinates, and imaged using a Canon EOS 600D DSLR camera with fixed parameters. Post-acquisition processing involved RAW image pre-processing, camera characterization (via polynomial regression mapping RAW RGB to CIE XYZ), and intentional introduction of WB estimation errors in terms of ΔCCT and ΔDuv. Chromatic adaptation to a D65 reference was conducted using the CAT16 model with fixed viewing parameters ($F = 1$, $L_A = 20$) [30]. For each reference illuminant, 32 deviated illuminants were generated via step-wise sampling across CCT and Duv axes, as well as their interaction directions.

For theatrical or colored lighting environments, 11 narrowband sources were generated using the LEDCube system by matching the selected chromaticity coordinates (see Figure 2). Twenty-four stimuli per source were generated via deviations along chroma (C), hue (h), and 45°/135° chromaticity axes.

The multiple illuminants (MI) scenarios were constructed also using ColorChecker chart under light sources with different CCTs and illuminance levels. All images underwent identical image pre-process and camera characterization procedure as in Figure 2. For the reference stimuli, the images were adapted to D65 using CAT16 with the correct illuminants before stitching. While for the testing stimuli, the images were first stitched and then performed CAT16 as a whole to simulated the estimated errors caused by WB algorithms assuming single-illuminant. The compositions included two different illuminated scenes with various area proportions (from 1:1 to 1:5), illuminance combinations (from 1:1 to 1:10), CCT combinations (four typical CCTs used in indoor lighting, including 3000 K, 4000 K, 5000 K, and 6500 K), and HDR scenes (the illuminance ratio of two sources reached 1:100, resulting a contrast ratio of 1:588 in the observed images).

Experimental Setups

All images were captured in a controlled darkroom environment utilizing an X-Rite ColorChecker Classic 24 chart, a Canon EOS 600D digital SLR camera, and an LEDCube light booth. The ColorChecker chart was mounted at a 45° angle within the LEDCube. The camera was mounted on a tripod and oriented perpendicular to the chart surface, with the optical axis centered on the target. Camera parameters were fixed throughout the data acquisition process: manual mode (M), 18 mm focal length, shutter speed 1/25 s, aperture f/5.6, ISO 100, and no flash or auto-WB. Images were captured using the Neutral Picture Style in three-frame burst mode per illumination setting.

The LEDCube light booth was employed to simulate the required illuminants by precisely mixing constituent single-color LEDs to replicate the spectral characteristics of various typical light sources. For single-illuminant scenarios, the target ColorChecker was imaged directly under each simulated light. In MI scenarios, the chart was captured sequentially under each light source within the combination. These separate, single-illuminant images were subsequently combined through image post-processing (stitching) to generate a composite image representing the dual-illuminant condition.

A calibrated LCD monitor (EIZO CS2740) was adopted in the visual acceptance assessment experiment. This display features a 4K resolution of 3840 × 2160 pixels, a contrast ratio of 1000:1, and a 99% Adobe RGB color gamut coverage. The monitor white point target was set to D65 with a maximum luminance of 100 cd/m^2. Participants were seated 40 cm from the screen during all assessment tasks.

Visual Assessment

A total of 27 observers (12 males, 15 females; mean age: 22.4 ± 3.1 years) participated in the PSI scenario. For CSI scenarios, 25 observers (10 males, 15 females; mean age: 23.7 ± 2.5 years) completed the canonical light experiment, while 24 observers (12 males, 12 females; mean age: 22.8 ± 3.9 years) participated in the colored light experiment. In the MI scenario, 25 observers (12 males, 13 females; mean age: 24.2 ± 2.8 years) were recruited. All participants were undergraduate or graduate students at Wuhan University, and all demonstrated normal color vision as confirmed by Ishihara screening.

Graphical user interfaces (GUIs) were developed to facilitate the visual acceptability assessment tasks, as shown in Fig. 3. For PSI scenarios, the interface presented pure color patches: the reference image (marked with an inverted triangle) was displayed alongside a color-shifted testing image on a black background (Fig. 3(a)). For CSI scenarios, ColorChecker chart images were observed and

assessed. Reference images were rendered with ground-truth illuminants (indicated by an inverted triangle), while testing images were rendered with deviated illuminants (Fig. 3(b)). For MI scenarios, reference images (rendered with ground-truth illuminants separately) were individually stitched to form a composite, then compared to testing images rendered under single-illuminant assumptions (Fig. 3(c)). All interfaces employed a consistent black background and standardized layout. In all scenarios, the left-right positioning of the reference and testing images, as well as the presentation order, were randomized to mitigate potential order effects.

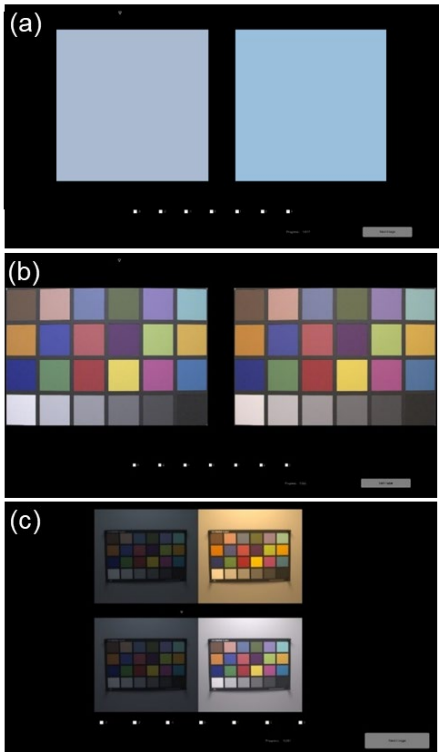

Figure 3. *Graphic user interfaces employed in the visual acceptance assessment tasks of (a) scenarios with pure colors and single illuminant, (b) scenarios with complex objects and single illuminant, and (c) multiple illuminant scenarios.*

Observers were instructed to undergo a two-minute dark adaptation period prior to the formal experiments. Subsequently, they evaluated the chromatic acceptability of the testing image relative to the reference image using a seven-point scale, where 1 ("completely unacceptable"), 2 ("unacceptable"), 3 ("barely unacceptable"), 4 ("neutral, i.e., the acceptability threshold"), 5 ("barely acceptable"), 6 ("acceptable"), and 7 ("completely acceptable"). Each observer participated in multiple sessions. For the PSI scenario, a total of 2,403 ratings were collected across all participants. For the CSI scenario, 525 ratings were obtained for canonical light experiments, and 286 ratings were collected for colored light experiments. In the MI scenario, a total of 251 ratings were recorded.

Results

Observer Variability

To evaluate both intra- and inter-observer consistency across experimental scenarios, we employed the standardized residual sum of squares (STRESS) metric, which ranges from 0 to 100, with lower values indicating higher consistency.

For the PSI scenario, the mean intra-observer STRESS (± standard deviation) was 21.99 (±5.57), indicating excellent repeatability. The corresponding inter-observer STRESS was slightly higher at 24.20 (±6.57). In the CSI scenario, the mean intra-observer STRESS was 22.34 (±7.75), with an inter-observer STRESS of 24.00 (±5.97) for canonical lights. Under colored light conditions, both intra-observer (27.05±9.44) and inter-observer (30.19±6.24) STRESS values increased, reflecting the perceptual complexity of colored illumination conditions.

The MI scenarios exhibited the highest overall STRESS values, with intra-observer STRESS at 28.73 (±7.72) and inter-observer STRESS at 30.23 (±8.45). Although elevated compared to single-illuminant scenarios, all values remained within the satisfactory range for observer variability [31,32], supporting the reliability of the assessment protocol.

Pure Colors with Single Illuminant (PSI) Scenarios

For the PSI scenarios, we compared several classical color difference models including DE2000 (DE_{00}), CIELAB (DE_{ab}), CIELUV (DE_{uv}), and CIECAM02-UCS ($DE_{CAM02UCS}$), using Pearson's correlation coefficient (r) to assess their ability to predict perceptual acceptability ratings. The absolute values of Pearson's r were used since observers' perceptual acceptability was negatively correlated with the magnitudes of color differences. As shown in Table 1, CIECAM02-UCS consistently demonstrated the highest predictive performance across all color centers (marked in bold). To match the observe environment, the color appearance computations were performed with Y_b and L_A set near zero, and the surround parameter set to "dark." Such a finding highlighted the significance of uniform color spaces and color appearance settings.

Table 1. Pearson's r performance of the color difference models for the Pure Colors with Single Illuminant (PSI) Scenarios.

Types of Color Centers	DE_{00}	DE_{ab}	DE_{uv}	$DE_{CAM02UCS}$
Neutral Colors	0.81	0.76	0.57	**0.90**
Low C* & Low L*	0.89	0.87	0.86	**0.91**
Low C* & Medium L*	0.81	0.84	0.89	**0.92**
Low C* & High L*	0.86	0.85	0.93	**0.91**
Medium C* & L*	0.78	0.90	0.89	**0.91**
Medium C* & High L*	0.84	0.87	0.84	**0.90**
High C*	0.76	0.79	0.83	**0.91**
Overall	0.80	0.75	0.73	**0.90**

In addition, by analyzing the magnitude of color differences across different color centers, it was also found that colors with higher lightness or chroma received higher acceptability ratings, indicating observers were more sensitive to low-lightness and low-chroma colors. Such effects were observed to be weaker in uniform color spaces like CIECAM02-UCS.

Complex Objects with Single Illuminant (CSI) Scenarios

Inspired by the findings above, we hypothesized that in a general scenario with various objects and colors, higher prediction performance could be achieved by incorporating the effects of different color attributes. Therefore, we introduced a weighted color difference metric defined as:

Table 2. Pearson's *r* performance of different color difference models (DE) and the weighted average color difference (WDE) for the Complex Objects with Single Illuminant (CSI) Scenarios.

Scenarios	Categories	DE_{00}	WDE_{00}	DE_{ab}	WDE_{ab}	DE_{uv}	WDE_{uv}	$DE_{CAM02UCS}$	$WDE_{CAM02UCS}$
Canonical light sources	Natural	0.88	**0.89**	0.91	0.90	0.89	**0.90**	0.88	**0.91**
	Domestic	0.90	**0.95**	0.92	**0.94**	0.90	**0.92**	0.93	**0.94**
	Shopping	0.89	**0.91**	0.90	0.90	0.90	0.89	0.92	0.92
	Working	0.87	**0.89**	0.90	**0.91**	0.88	**0.91**	0.89	**0.90**
Colored light sources	Medium C*	0.75	**0.93**	0.74	**0.90**	0.68	**0.85**	0.77	**0.89**
	High C*	0.74	**0.95**	0.78	**0.93**	0.71	**0.90**	0.78	**0.93**

$$\mathrm{WDE} = \sum_{i=1}^{n} (w_L \cdot L_i^* + w_a \cdot a_i^* + w_b \cdot b_i^*) \cdot DE_i \qquad (1)$$

Where WDE represents the weighted average color difference between the original colors and deviated colors for multiple-colored images and n represents the number of colors in the image. In this study, n = 24, representing 24 ColorChecker. And w_L, w_a, w_b represent the weights for the L*, a*, b* values of colors in the image, indicating the contribution of different colors to the subjective acceptability of the overall image. These weights were optimized to achieve the strongest correlation between DE_{MC} and observers' acceptance ratings. DE_i represents the color difference between the original color and deviated color in the image.

The results were shown in Table 2, cases where our WDE metrics showed improvement over traditional models were marked in bold. It can be observed that without WDE metric, CIECAM02-UCS again showed the best performance in colored light scenarios, though the overall correlation was evidently lower than in canonical light conditions. On the other hand, the WDE metric we proposed could significantly improve all models' predictive performance, especially for DE2000 formula.

Multiple Illuminant (MI) Scenarios

The investigation into multi-illuminant scenarios revealed systematic patterns in perceptual acceptability. For area proportion effects, images with substantial chromatic deviations consistently received "unacceptable" ratings (lower than 4) across all area ratios. In addition, regions with larger proportions exhibited stronger impacts on acceptability, particularly under extreme chromatic deviation. Similarly, illuminance ratio variations revealed that most configurations remained "unacceptable". Notably, the region with higher illuminance had a greater influence on observer ratings.

Results of various CCT combinations demonstrated that acceptability ratings improved as the CCTs of the two illuminants converged, and intermediate CCT renderings consistently resulted in higher acceptability scores. In HDR scenarios, only images with a 1:100 contrast ratio rendered using 6500 K illuminant achieved "acceptable" ratings. This may be due to observers focusing on the high-luminance region, which displayed no color cast, while the lower-luminance region received less attention

Based on these findings, we propose a weighted color difference metric incorporating spatial and illuminance factors:

$$DE_{MI} = w_{left} * DE_{left} + w_{right} * DE_{right} \qquad (2-1)$$

$$w_{left} = \left(m_{area} \cdot A_{left} + d_{area}\right) \cdot \left(m_E \cdot E_{left} + d_E\right) \qquad (2-2)$$

$$w_{right} = \left(m_{area} \cdot A_{right} + d_{area}\right) \cdot \left(m_E \cdot E_{right} + d_E\right) \qquad (2-3)$$

Where DE_{MI} represents the weighted color difference in multiple illumination scenarios, w_{left} and w_{right} represent the weights considering the area ratio and illuminance ratio between the left illumination scenario and the right illumination scenario. DE_{left} and DE_{right} represent the colour difference between the original colors and deviated colors for the left and right illumination scenarios, respectively. $A_{left}{:}A_{right}$ is the area ratio and $E_{left}{:}E_{right}$ is the illuminance ratio between left and right illumination scenarios. Finally, m_{area}, d_{area}, m_E, and d_E were fitted parameters that optimized to achieve the strongest correlation between DE_{MIS} and observers' acceptance ratings.

Although models' performance was generally suboptimal in specific MI subcategories, as shown in Table 3, our proposed method consistently yielded strong predictive power in the overall results. The disparity between subgroup and overall outcomes likely stems from the limited sample sizes in the subgroups.

Table 3. Pearson's *r* performance of the weighted color difference models in the Multiple Illuminants (MI) scenarios.

Categories	DE_{00}	DE_{ab}	DE_{uv}	$DE_{CAM02UCS}$
Area ratio	0.40	0.51	0.72	0.39
Illuminance ratio	0.48	0.45	0.50	0.54
Interactions	0.50	0.55	0.59	0.53
CCTs	0.61	0.72	0.80	0.60
HDR (3000K+6500K)	0.94	0.95	0.96	0.96
HDR (6500K)	0.42	0.47	0.46	0.40
Overall	**0.91**	**0.89**	**0.88**	**0.83**

Discussion

Our findings reveal fundamental perceptual hierarchies in color acceptability assessment. Notably, observers demonstrated heightened sensitivity to chromatic deviations in low-lightness and low-chroma regions, suggesting that existing color difference models may underestimate perceptual significance in these areas while uniform color spaces like CIECAM02-UCS performed relatively better. Based on these findings, a weighted metric based on color appearance attributes was introduced and applied on existing color difference models to better reflect observed acceptability ratings, particularly for subtle shifts in achromatic and near-neutral regions where human vision exhibits higher sensitivity.

Spatial and illuminance factors critically modulated acceptability judgments in multi-illuminant scenarios. When substantial chromatic deviations occurred, neither area proportion nor illuminance ratio variations could yield ratings beyond the acceptability threshold. Crucially, larger and brighter image regions exerted greater influence on observer responses—a dominance effect amplified under extreme deviations. To quantify these interactions, we propose an area-illuminance weighted metric incorporating spatial and illuminance weighting coefficients. While this metric captured key empirical trends and achieved high predictive accuracy in overall results, it remained limited in certain subcategories and required further validation across broader illumination configurations.

Evaluation of color difference models yielded context-dependent performance. For instance, DE2000 generally produced strongest correlations in scenarios with complex objects while its performance was found to be worse than CIECAM02-UCS in PSI scenes. On the other hand, CIECAM02-UCS demonstrated strong

correlation with observer data in single-illuminant scenarios (Pearson's $r < -0.89$ for both pure colors and complex objects contexts), but its efficacy diminished in the subcategories of multi-illuminant scenarios (Pearson's r ranged from -0.39 to -0.96), revealing gaps in modeling spatially ununiform illumination conditions. These limitations should be contextualized by our selective assessment of classical color difference models, excluding parametric effects and newer metrics. Future research should incorporate expanded model comparisons including more recent work [30,33–35], parametric effect analysis across color difference models [36–38], and integration of metrics such as angular errors [7,12]for direct comparison.

Conclusions

This study provides new insights into perceptual white balance (WB) evaluation by shifting the focus from traditional illuminant-chromaticity-based metrics to observer-based acceptability in diverse lighting environments. Through a rigorously constructed Multiple Illumination Scenarios (MIS) dataset and extensive psychophysical experiments, we revealed that perceptual judgments are strongly modulated by color appearance attributes and spatial illumination compositions. The proposed weighted metrics based on both color appearance attributes and spatial-illuminance factors could significantly enhance prediction accuracy of existing color difference models, particularly in multi-illuminant scenarios where conventional methods underperform. By incorporating more sophisticated and realistic datasets that reflect the diversity of real-world scenarios and integrating alternative color difference models and WB error evaluation metrics, we hope to further advance the research of WB error acceptability prediction under complex lighting environments.

References

1. F. Wang, W. Wang, and Z. Qiu, "Color correction algorithm for color constancy finite dimensional linear model under complex illumination," Pattern Recognit. Lett. **116**, 233–237 (2018).
2. R. Zakizadeh and G. D. Finlayson, "A Psychophysical Analysis of Illuminant Estimation Algorithms," Color Imaging Conf. **25**(1), 70–75 (2017).
3. A. Gijsenij, T. Gevers, and J. van de Weijer, "Computational Color Constancy: Survey and Experiments," IEEE Trans. Image Process. **20**(9), 2475–2489 (2011).
4. M. Afifi and M. S. Brown, "Deep White-Balance Editing," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, 2020), pp. 1394–1403.
5. D. Serrano-Lozano, A. Arora, L. Herranz, K. G. Derpanis, M. S. Brown, and J. Vazquez-Corral, "Revisiting Image Fusion for Multi-Illuminant White-Balance Correction," (2025).
6. M. Afifi, M. A. Brubaker, and M. S. Brown, "Auto White-Balance Correction for Mixed-Illuminant Scenes," in *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)* (IEEE, 2022), pp. 934–943.
7. D. A. Forsyth, "A novel algorithm for color constancy," Int. J. Comput. Vis. **5**(1), 5–35 (1990).
8. E. H. Land, "The Retinex Theory of Color Vision," Sci. Am. **237**(6), 108–128 (1977).
9. Marc Ebner, "Algorithms for Color Constancy under Nonuniform Illumination," in *Color Constancy* (John Wiley & Sons, Ltd, 2006), pp. 143–191.
10. O. Sidorov, "Conditional GANs for Multi-Illuminant Color Constancy: Revolution or yet Another Approach?," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (2019), pp. 1748–1758.
11. Marc Ebner, "Evaluation of Algorithms," in *Color Constancy* (John Wiley & Sons, Ltd, 2006), pp. 275–301.
12. G. Finlayson and R. Zakizadeh, "Reproduction Angular Error: An Improved Performance Metric for Illuminant Estimation," in *Proceedings of the British Machine Vision Conference 2014* (British Machine Vision Association, 2014), p. 70.1-70.11.
13. N. Banić and S. Lončarić, "A Perceptual Measure of Illumination Estimation Error:," in *Proceedings of the 10th International Conference on Computer Vision Theory and Applications* (SCITEPRESS - Science and and Technology Publications, 2015), pp. 136–143.
14. G. D. Finlayson, B. V. Funt, and K. Barnard, "Color constancy under varying illumination," in *Proceedings of IEEE International Conference on Computer Vision* (1995), pp. 720–725.
15. G. D. Finlayson and R. Zakizadeh, "Reproduction angular error: An improved performance metric for illuminant estimation: 25th British Machine Vision Conference, BMVC 2014," Proc. Br. Mach. Vis. Conf. 2014 (2014).
16. Q. Zhang and R. L. Canosa, "A comparison of histogram distance metrics for content-based image retrieval," in *Imaging and Multimedia Analytics in a Web and Mobile World 2014* (SPIE, 2014), **9027**, pp. 154–162.
17. M. R. Luo, G. Cui, and B. Rigg, "The development of the CIE 2000 colour-difference formula: CIEDE2000," Color Res. Appl. **26**(5), 340–350 (2001).
18. M. R. Luo, G. Cui, and C. Li, "Uniform colour spaces based on CIECAM02 colour appearance model," Color Res. Appl. **31**(4), 320–330 (2006).
19. S.-S. Guan and M. R. Luo, "A colour-difference formula for assessing large colour differences," Color Res. Appl. **24**(5), 344–355 (1999).
20. G. Cui, M. Luo, B. Rigg, G. Roesler, and K. Witt, "Uniform colour spaces based on the DIN99 colour-difference formula," Color Res. Appl. **27**, 282–290 (2002).
21. CIE 15:2004 Colorimetry, *CIE 15:2004 Colorimetry* (2004).
22. M. R. Luo and B. Rigg, "Chromaticity-discrimination ellipses for surface colours," Color Res. Appl. **11**(1), 25–42 (1986).
23. D. T. Lindsey and A. G. Wee, "Perceptibility and acceptability of CIELAB color differences in computer-simulated teeth," J. Dent. **35**(7), 593–599 (2007).
24. R. D. Paravina, M. M. Pérez, and R. Ghinea, "Acceptability and perceptibility thresholds in dentistry: A comprehensive review of clinical and research applications," J. Esthet. Restor. Dent. **31**(2), 103–112 (2019).
25. S. Yue and M. Wei, "Color constancy from a pure color view," JOSA A **40**(3), 602–610 (2023).

26. L. Chen, M. R. Luo, M. Wei, M. R. Luo, and M. Wei, "Large Size of Color Constancy: Enhancing Pure Color Image Illuminant Estimation with Kolmogorov-Arnold Networks," Color Imaging Conf. **32**, 95–100 (2024).
27. P. Balakrishnan, D. Dumortier, P. Kenny, M. Maskarenj, C. Pierson, A. Thorseth, P. Xue, M. Knoop, J. Hernández-Andrés, and A. K. Diakite-Kortlever, "SKYSPECTRA: AN OPENSOURCE DATA PACKAGE OF WORLDWIDE SPECTRAL DAYLIGHT," CIE X0502023 Proc. 30th Sess. CIE Ljubl. Slov. Sept. 15 – 23 2023 Vol. 1 (2023).
28. Z. Huang, W. Chen, Q. Liu, Y. Wang, M. R. Pointer, Y. Liu, and J. Liang, "Towards an optimum colour preference metric for white light sources: a comprehensive investigation based on empirical data," Opt. Express **29**(5), 6302–6319 (2021).
29. Q. Liu, Y. Liu, M. R. Pointer, Z. Huang, X. Wu, Z. Chen, and M. R. Luo, "Color discrimination metric based on the neutrality of lighting and hue transposition quantification," Opt. Lett. **45**(21), 6062–6065 (2020).
30. C. Li, Z. Li, Z. Wang, Y. Xu, M. R. Luo, G. Cui, M. Melgosa, M. H. Brill, and M. Pointer, "Comprehensive color solutions: CAM16, CAT16, and CAM16-UCS," Color Res. Appl. **42**(6), 703–718 (2017).
31. H. Wang, G. Cui, M. R. Luo, and H. Xu, "Evaluation of colour-difference formulae for different colour-difference magnitudes," Color Res. Appl. **37**(5), 316–325 (2012).
32. M. Melgosa, P. A. García, L. Gómez-Robledo, R. Shamey, D. Hinks, G. Cui, and M. R. Luo, "Notes on the application of the standardized residual sum of squares index for the assessment of intra- and inter-observer variability in color-difference experiments," JOSA A **28**(5), 949–953 (2011).
33. M. Li and M. R. Luo, "Simple color appearance model (sCAM) based on simple uniform color space (sUCS)," Opt. Express **32**(3), 3100 (2024).
34. S. Hermans, K. a. G. Smet, and P. Hanselaer, "Color appearance model for self-luminous stimuli," JOSA A **35**(12), 2000–2009 (2018).
35. M. Safdar, J. Y. Hardeberg, and M. R. Luo, "ZCAM, a colour appearance model based on a high dynamic range uniform colour space," Opt. Express **29**(4), 6036–6052 (2021).
36. M. Luo, Q. Xu, M. Pointer, M. Melgosa, G. Cui, li Changjun, K. Xiao, and M. Huang, "A comprehensive test of colour-difference formulae and uniform colour spaces using available visual datasets," Color Res. Appl. **48**, (2023).
37. Q. Xu, B. Zhao, G. Cui, and M. Luo, "Testing uniform colour spaces using colour differences of a wide colour gamut," Opt. Express **29**, (2021).
38. Q. Xu, K. Shi, and M. Luo, "Parametric effects in color-difference evaluation," Opt. Express **30**, (2022).