

# Visualizing Perceptual Differences in White Color Constancy

Marco Buzzelli

Department of Informatics Systems and Communication, University of Milano – Bicocca, 20126 Milan, Italy

E-mail: marco.buzzelli@unimib.it

---

**Abstract.** Algorithms for computational color constancy are usually compared in terms of the angular error between ground truth and estimated illuminant. Despite its wide adoption, there exists no well-defined consensus on acceptability and/or noticeability thresholds in angular errors. One of the main reasons for this lack of consensus is that angular error weighs all hues equally by performing the comparison in a non-perceptual color space, whereas the sensitivity of the human visual system is known to vary depending on the chromaticity. We therefore propose a visualization strategy that presents simultaneously the angular error (preserved due to its wide adoption in the field), and a perceptual error (to convey information about its actual perceived impact). This is achieved by exploiting the angle-retaining chromaticity diagram, which shows errors in chromaticities while encoding RGB angular distances as 2D Euclidean distances, and by embedding contour lines of perceptual color differences at standard predefined thresholds. Example applications are shown for different color constancy methods on two imaging devices. © 2023 Society for Imaging Science and Technology.  
[DOI: 10.2352/J.ImagingSci.Technol.2023.67.5.050404]

---

## 1. INTRODUCTION

Human color constancy has been defined as the ability to recognize the colors of objects independent of the characteristics of the light source [1]. Computational color constancy, hereafter referred to as “color constancy” for brevity, is a technique developed in digital imaging to emulate the process of human color constancy within the context of digital sensors. Its implementation in consumer devices is often referred to as “Automatic White Balance”, in contrast to “Manual White Balance” where the user is in charge of removing (or changing) the color cast from the acquired image.

Color constancy is typically addressed as a two-stage process, with illuminant estimation followed by illuminant correction. The first stage aims at characterizing the illuminant source, and it is the main focus of most algorithms for color constancy. The second stage employs a mechanism for chromatic adaptation to correct the image. The most common and simple solution for illuminant correction consists in applying to tristimulus data a diagonal transformation matrix, referred to as a “wrong von Kries transform” [2]. Color constancy is commonly evaluated in terms of angular errors [3–5]. The recovery angular error [4] treats the ground truth illuminant and the estimated

illuminant as vectors in RGB space, and measures the angle between such vectors, thus comparing only the chromaticities and ignoring their absolute intensities. The reproduction angular error [5] simulates the effect of a possibly-wrong estimation on the correction of a neutral surface; the ground truth illuminant (representing the appearance of a neutral surface) is normalized by the estimation, applying a correction through a wrong von Kries transform. The resulting vector is then compared in terms of its angular distance with the vector of perfect white, defined by  $RGB = [1, 1, 1]$ . Despite the wide adoption of angular error metrics for the evaluation and comparison of color constancy algorithms, there exists no consensus about angular thresholds for acceptability and noticeability. After informal experiments first conducted in 2005, Finlayson et al. [6] considered a  $1^\circ$  perturbation of the image illuminant in any given color direction as not visibly noticeable, while a  $3^\circ$  perturbation was found to be noticeable but “generally acceptable”, as later supported by Fredembach et al. [7] and adopted as a threshold in the comparison of color constancy methods [8]. A  $5^\circ$  perturbation was defined as “generally acceptable, but unacceptable for some images”, thus manifesting the impossibility to assign a unique threshold for acceptability, even within the same experiment. Similarly, in 2006, Hordley [9] conducted another informal analysis, suggesting an angular error of  $2^\circ$  to be acceptable for color constancy. Rather than proposing an absolute threshold, Gijsenij et al. [10, 11] addressed the concept of noticeability in relative terms, determining that the difference in terms of angular error between two methods should be at least 0.06 times the maximum of the two errors, in order for it to be noticeable. This idea was later adopted for method comparison to determine the significance of angular error differences [12]. Gijsenij et al. [10, 11] also conducted a broader analysis of alternative metrics for error evaluation of color constancy, proposing a weighted Euclidean distance, and a corresponding coefficient for the threshold of relative noticeability. Other authors also considered alternative metrics and domains for the definition of color thresholds. Hordley [9] reported a CIELab error of 1 as a just noticeable threshold for two colors viewed side by side in isolation, and a CIELab error of 6 as an acceptable threshold for the assessment of complex scenes.

An intrinsic limitation of defining acceptability and noticeability in RGB space is that such representation is not perceptually uniform, whether it is assumed to be sRGB or a device-specific raw-RGB. A possible solution to the

---

Received May 10, 2023; accepted for publication Aug. 9, 2023; published online Sept. 6, 2023. Associate Editor: Sauek Morillas.

1062-3701/2023/67(5)/050404/6/\$25.00

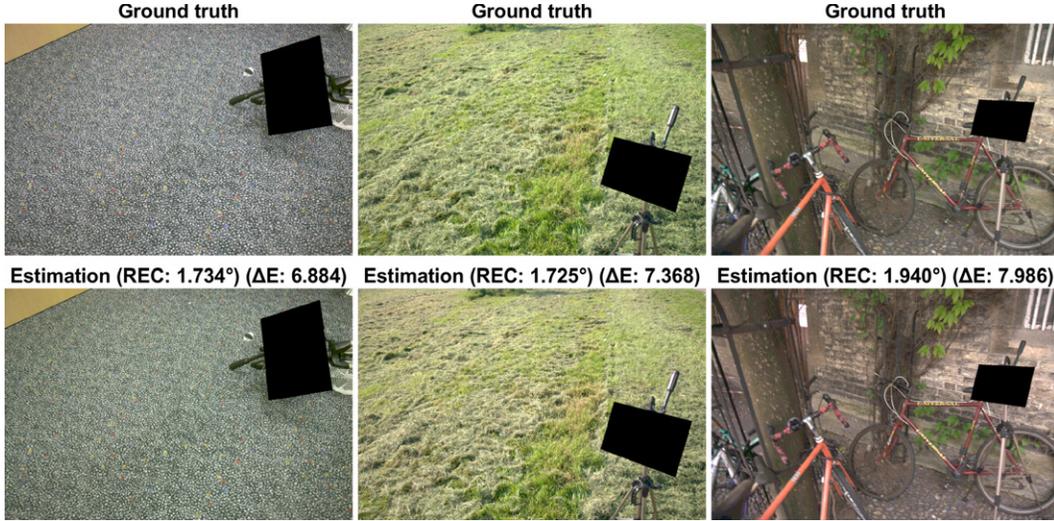


Figure 1. Sample corrections from FC4 [14]: recovery angular error is below the  $2^\circ$  acceptability threshold, but  $\Delta E * ab$  is above the acceptability threshold of 6.

problem lies in measuring the error through perceptual difference metrics, such as the CIE76  $\Delta E * ab$  [13], based on the Euclidean distance between two colors in CIELab color space. One advantage of this solution is the availability of de facto standard thresholds, such as the JND (Just Noticeable Difference) set to 2.3, or the acceptability in complex scenes defined by Hordley to be 6. This approach has, however, two main drawbacks: first, it ignores the reality that the majority of methods comparison in the domain of color constancy is, in fact, performed with angular error metrics, thus raising a retro-compatibility issue. Secondly, any single-value metric will, by definition, remove any information about the chromaticity of the estimation error, thus depriving of an important clue for the analysis of a method's faults (this limit is shared with the traditional angular error). Furthermore, there exists a disagreement between acceptability thresholds defined in angular error and  $\Delta E * ab$ , as manifest in Figure 1: here, three example corrections from the fully convolutional color constancy with confidence-weighted pooling (FC4) method [14] are visualized, together with the corresponding recovery angular error (always under  $2^\circ$ , considered acceptable), and the corresponding  $\Delta E * ab$  (always above 6, considered unacceptable).

For these reasons, we propose a solution for the visualization of errors in the domain of color constancy, based on embedding contour lines of perceptual color differences at predefined thresholds into a chromaticity diagram that encodes RGB angular distances as 2-dimensional Euclidean distances, called Angle-Retaining Chromaticity (ARC) [15, 16]. First, using a chromaticity diagram enables the visualization of errors' chromaticities, thus providing an intrinsically richer representation of the errors. Secondly, with this configuration, the errors can be visualized both in terms of the commonly used angular error (represented by the Euclidean distance in 2D space) and in terms of the more

informative CIE76  $\Delta E * ab$  (with thresholds represented by contour lines).

## 2. BACKGROUND ON COLOR ERROR REPRESENTATION

Let  $G = \{g_R, g_G, g_B\}$  be a ground truth illuminant in device-specific raw-RGB, and let  $E = \{e_R, e_G, e_B\}$  be an estimated illuminant in the same color space. The angular distance between illuminants  $G$  and  $E$  is the recovery angular error [4]. Alternatively, the reproduction angular error [5] is obtained through the reproduction vector  $\mathbf{R}$ :

$$\mathbf{R} = \frac{G}{E} = \left\{ \frac{g_R}{e_R}, \frac{g_G}{e_G}, \frac{g_B}{e_B} \right\}. \quad (1)$$

The angular distance between the reproduction vector  $\mathbf{R}$  and a neutral surface  $W = \{1, 1, 1\}$  is, by definition, the reproduction angular error.

The ARC diagram [15, 16] is a general-purpose 2D representation of color information, which may be obtained from any given RGB vector  $\{\rho_R, \rho_G, \rho_B\}$  as follows:

$$\alpha_A = \arctan 2 \left( \sqrt{3} (\rho_G - \rho_B), 2\rho_R - \rho_G - \rho_B \right) \quad (2)$$

$$\alpha_R = \arccos \left( \frac{\rho_R + \rho_G + \rho_B}{\sqrt{3} \sqrt{\rho_R^2 + \rho_G^2 + \rho_B^2}} \right) \quad (3)$$

$$\alpha_Z = \sqrt{\rho_R^2 + \rho_G^2 + \rho_B^2}. \quad (4)$$

Alternatively, the polar chromaticity coordinates  $\{\alpha_A, \alpha_R\}$  may be represented in Cartesian form  $\{\alpha_X, \alpha_Y\}$  for visualization purposes. When converting a reproduction vector  $\mathbf{R}$  into ARC, each pair of ground truth and estimated illuminants becomes a single two-dimensional point, whose Euclidean distance from the diagram center corresponds exactly to the reproduction error.

### 3. PROPOSED METHODOLOGY FOR VISUALIZATION

Our goal is to assign a perceptual error to every point in the ARC diagram and to eventually visualize this information with contour lines. Such error is always computed between the reproduction vector  $\mathbf{R}$  and the white vector  $\mathbf{W}$ , in line with the computation of the reproduction angular error. This strategy entails a reduction of dimensionality that enables data visualization since displaying perceptual errors for all possible illuminant pairs would not be possible. As each illuminant is represented by at least two chromaticity coordinates, in fact, this would lead to four independent variables and one dependent variable (the error value). By contrast, the advantage of always referencing a single point for comparison ( $W$ ) is that the only independent variable is the two-dimensional ratio vector in chromaticity, thus allowing for error visualization.

Let the coordinate system of the reproduction vector  $\mathbf{R}$  be called “reproduction- $\text{RGB}$ ”, which differs from raw- $\text{RGB}$  as explained in the following. The procedure used to compute the perceptual error for  $\mathbf{R}$  is delineated here, which starts by bringing the reproduction vector back into raw- $\text{RGB}$  color space.

#### 3.1 Color Space Conversion: Reproduction- $\text{RGB}$ to raw- $\text{RGB}$

When the reproduction vector  $\mathbf{R}$  is computed according to (Eq. (1)), it is brought outside of the original raw- $\text{RGB}$  space. This can be shown by observing that, for example, the white point  $W = \{1, 1, 1\}$  in this  $\text{RGB}$  space will not be rendered as white using the camera-specific raw-to- $\text{CIElab}$  transformation. The necessary conversion is therefore one that first transforms, in our example, the reproduction- $\text{RGB}$   $W$  into a raw- $\text{RGB}$  that would, in turn, render as a white  $\text{CIElab}$ . Adhering to the von-Kries-like model that underlies the reproduction error metric, we achieve this goal by multiplying  $\mathbf{R}$  by a vector  $\mathbf{D}$ , which simulates the application of an illuminant to a surface. Vector  $\mathbf{D}$  is computed as:

$$D = xyz2raw(lab2xyz([100, 0, 0])), \quad (5)$$

where  $lab2xyz(\cdot)$  is a standard transformation, and  $xyz2raw(\cdot)$  is a device-specific transformation matrix.

#### 3.2 Color Space Conversion: Raw- $\text{RGB}$ to $\text{XYZ}$

The conversion from raw to  $\text{XYZ}$  is obtained with a transformation matrix that is specific for the individual camera. This implies that every device will produce its own device-specific set of contour lines.

#### 3.3 $Y$ Normalization

The  $\text{XYZ}$ -encoded reproduction vector is normalized (i.e. divided) by its  $Y$  component, to discard the intensity information. Color constancy is, in fact, traditionally evaluated by ignoring intensity, to the extent that some algorithm implementations directly produce normalized  $\text{RGB}$  estimations.

#### 3.4 Color Space Conversion: Normalized $\text{XYZ}$ to $\text{CIElab}$

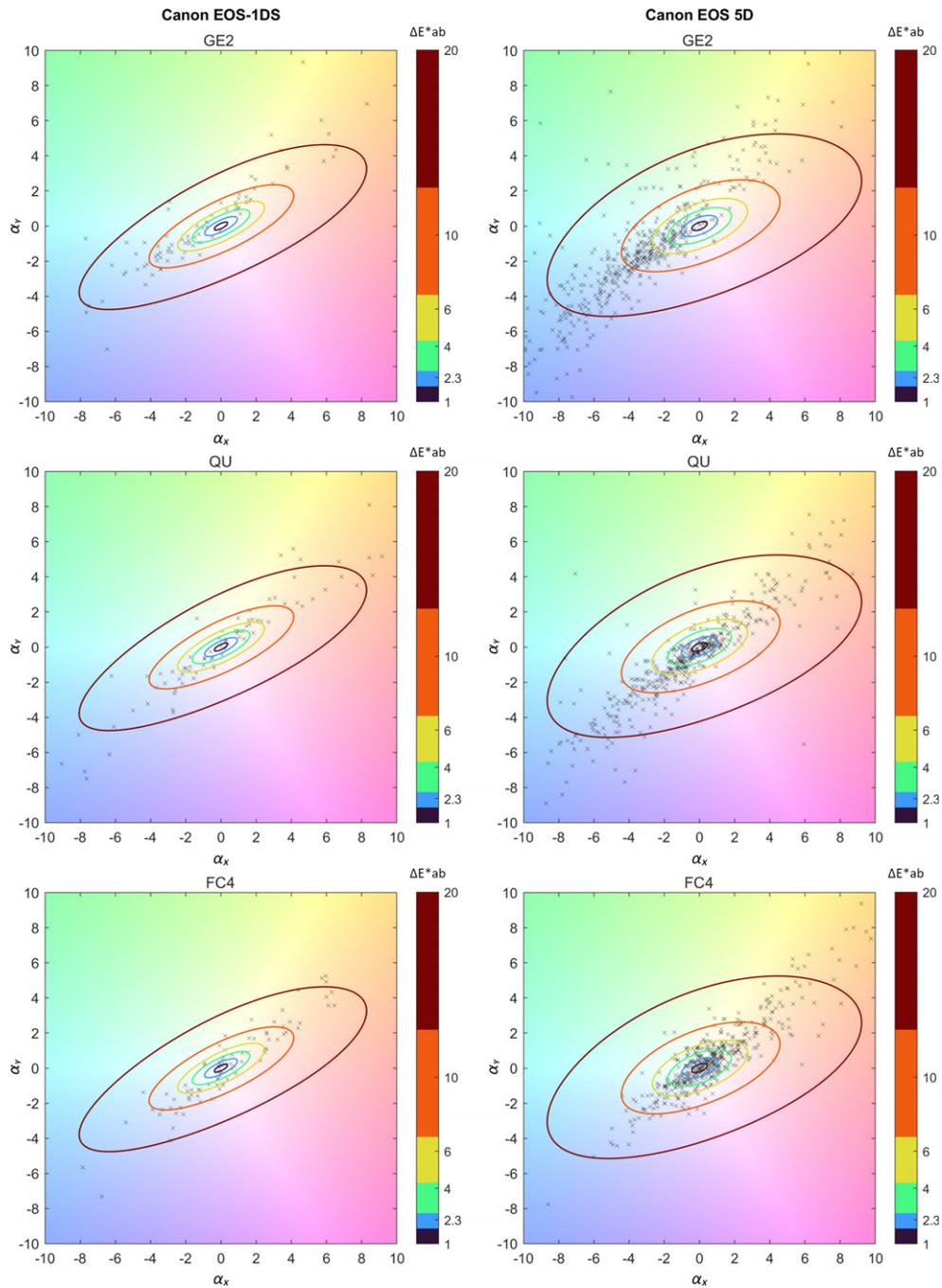
The final transformation uses a standard conversion from  $\text{XYZ}$  to  $\text{CIElab}$ , assuming  $D65$  as the reference white. The rationale is that we are observing a picture in which the illuminant has been corrected using a given estimation, and our observation of the corrected picture takes place under  $D65$ .

Once the reproduction vector  $\mathbf{R}$  is converted into  $\text{CIElab}$ , the same procedure is applied to the white  $\text{RGB}$  vector  $W = \{1, 1, 1\}$ , and the  $\Delta E * ab$  distance is computed between the two. This operation is repeated for all points in the chromaticity diagram, thus producing a dense map of perceptual distances, which can be visualized as overlaid contour lines.

### 4. RESULTING VISUALIZATIONS

In this section, we apply the proposed visualization to images from the ColorChecker dataset [17]. This dataset is composed of 86 images acquired with a Canon EOS-1DS camera, and 482 images acquired with a Canon EOS 5D camera. Every image includes a Macbeth ColorChecker target within the scene. From this, the brightest of the six achromatic patches with no oversaturated channels is used to determine the ground truth illuminant for the color constancy task, following Hemrit et al. [18]. For both devices in the dataset, we reference the corresponding model-specific raw-to- $\text{XYZ}$  transformation matrices from the DCRAW software (<https://www.dechifro.org/dcrow/>), since no calibration data is available and all images include a color target for device-specific illuminant annotation, no absolute reference is provided, thus preventing the definition of a proper color profile.

We compare estimations from three methods for computational color constancy, selected as representatives of different categories; the second-order Grey Edge (GE2) in its implementation by van de Weijer et al. [19] as an instance of a statistics-based algorithm, Quasi-Unsupervised color constancy (QU) by Bianco et al. [20] as an instance of semi-supervised machine learning, and FC4 by Hu et al. [14] as an instance of fully-supervised machine learning. The learning-based methods QU and FC4 are designed for full white balancing, i.e. they are trained to identify neutral regions in the scene, and to consequently correct the image so that these regions are rendered as gray. GE, on the other hand, is purely an illuminant estimation method, with no assumptions on how to perform the final correction. Additionally, we note that this selection of methods is to be intended as a use case for the general application of our visualization methodology to any single-illuminant algorithm for the color constancy of white. The relationship between angular errors and  $\Delta E * ab$ , visually presented in Fig. 1, is synthesized in Table I. For all combinations of cameras and methods, we report aggregate error statistics as well as the correlation between angular error (recovery or reproduction) and  $\Delta E * ab$ . This incomplete correlation is later explored thanks to our visualization.



**Figure 2.** Visualization of color constancy errors for three methods (GE2, QU, FC4) on two cameras of the ColorChecker dataset. Crosses represent reproduction errors based on single-patch ground truth data. Contour lines represent  $\Delta E^*ab$  thresholds at steps indicated in the color bars. The dimensionless axes of the ARC diagram represent RGB angular distances as two-dimensional Euclidean distances.

Our proposed visualization is presented in the chromaticity diagrams of Figure 2. Here, Euclidean distances from the center correspond to the reproduction angular error expressed in degrees and can be visually judged from the diagram’s own frame of reference. The rendered  $\Delta E^*ab$  contour lines visually highlight how the sensitivity of the human visual system is not fully correlated to the absolute angular error, depending instead on the hue of the error

itself. Specifically, errors distributed on the correlated color temperatures of a black body radiator, moving from blue to yellow/red, are less perceivable than errors distributed on the orthogonal axis, moving from green to magenta. Both cameras from the ColorChecker dataset lead to similar conclusions, although a difference in sensitivity on the green-magenta axis can be clearly observed. Every cross ( $\times$ ) in the diagram is a reproduction vector computed

**Table I.** Summary statistics (mean and 99th percentile) of analyzed color constancy methods on the two cameras of the ColorChecker dataset [17, 18]. Pearson correlation between angular errors and  $\Delta E^*ab$  is also reported.

Camera	Method	Rec. Err. (°)		Rep. Err (°)		$\Delta E^*ab$		Pearson corr.	
		Mean	99th p.	Mean	99th p.	Mean	99th p.	Rec./ $\Delta E$	Rep./ $\Delta E$
Canon EOS-1Ds	GE2 [19]	3.754	14.668	4.464	16.729	11.280	40.103	0.9743	0.9533
	QU [20]	3.644	13.517	4.496	18.078	10.439	39.852	0.9858	0.9796
	FC4 [14]	2.332	8.356	2.901	9.871	7.136	23.567	0.9763	0.9580
Canon EOS-5D	GE2 [19]	4.127	14.310	5.242	18.227	11.934	41.426	0.9788	0.9623
	QU [20]	2.853	13.915	3.678	18.324	8.257	38.007	0.9816	0.9570
	FC4 [14]	1.849	8.773	2.396	12.241	5.543	23.679	0.9864	0.9620

according to Eq. (1) using the single gray-patch-based ground truth for a given image, and the corresponding illuminant estimation obtained from a given method. The three analyzed methods produce errors distributed in different regions of the chromaticity diagram, with GE2 and QU exhibiting a stronger bias towards blue overcorrections, and FC4 towards yellow overcorrections. In general, they all display an error distribution that roughly follows the blue to yellow/red axis of a black body radiator. While this may be expected of data-driven methods such as QU and FC4, which tend to inherit the bias of their corresponding training datasets, it is a surprising result for the data-free GE2 method. A possible explanation lies in the inherent limitation of annotating a single illuminant in what is actually a multiple-illuminant scenario [21] due, for example, to mutual surface inter-reflections or the coexistence of sun and shadow areas. In this case, the chromaticity of the error is potentially revealing non-annotated illuminant information. This interpretation could be further tested by extending the evaluation to images acquired under non-blackbody illuminants, a feature that is known to be lacking from many of the existing datasets for color constancy [3].

## 5. CONCLUSIONS AND FUTURE WORKS

We have presented a visualization technique for color constancy errors that visualizes, at the same time, the traditionally-used angular error, its chromaticity, and its perceivability. Our visualization has been applied to three color constancy methods with different levels of complexity and accuracy, on two cameras of the popular ColorChecker dataset, providing a broad overview of its applicability. For future development, we consider expanding this analysis to a wider range of methods, datasets, and perceptual metrics. An inherent difficulty in assigning a perceivability score to color constancy errors derives from the necessity to evaluate the whole image in context, as opposed to only comparing colors “in a void”, which poses an additional challenge in terms of data visualization.

## ACKNOWLEDGMENT

This work was partially supported by the MUR under the grant “Dipartimenti di Eccellenza 2023–2027” of the Department of Informatics, Systems and Communication of the University of Milano-Bicocca, Italy. This work was partially funded under the National Recovery and Resilience Plan (NRRP), Mission 4 Component 2 Investment 1.3 - Call for tender No. 341 of 15 March 2022 of Italian Ministry of University and Research funded by the European Union – NextGenerationEU; Award Number: Project code PE00000003, Concession Decree No. 1550 of 11 October 2022 adopted by the Italian Ministry of University and Research, CUP D93C22000890001, Project title “ON Foods - Research and innovation network on food and nutrition Sustainability, Safety and Security – Working ON Foods”.

## REFERENCES

- M. Buzzelli, J. van de Weijer, and R. Schettini, “Learning illuminant estimation from object recognition,” *2018 25th IEEE Int’l. Conf. on Image Processing (ICIP)* (IEEE, Piscataway, NJ, 2018), pp. 3234–3238.
- J. von Kries, *Theoretische studien über die umstimmung des sehorgans* (Festschrift der Albrecht-Ludwigs-Universität, 1902), pp. 145–158.
- M. Buzzelli, S. Zini, S. Bianco, G. Ciocca, R. Schettini, and M. K. Tchobanou, “Analysis of biases in automatic white balance datasets and methods,” *Color Res. Appl.* **48**, 40–62 (2023).
- S. D. Hordley and G. D. Finlayson, “Reevaluation of color constancy algorithm performance,” *J. Opt. Soc. Am. A* **23**, 1008–1020 (2006).
- G. D. Finlayson and R. Zakizadeh, “Reproduction angular error: An improved performance metric for illuminant estimation,” *Perception* **310**, 1–26 (2014).
- G. D. Finlayson, S. D. Hordley, and P. Morovic, “Colour constancy using the chromagenic constraint,” *2005 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR’05)* (IEEE, Piscataway, NJ, 2005), 1, pp. 1079–1086.
- C. Fredembach and G. Finlayson, “The bright-chromagenic algorithm for illuminant estimation,” *J. Imaging Sci. Technol.* **52**, 040906-1 (2008).
- S. Bianco, C. Cusano, and R. Schettini, “Single and multiple illuminant estimation using convolutional neural networks,” *IEEE Trans. Image Process.* **26**, 4347–4362 (2017).
- S. D. Hordley, “Scene illuminant estimation: past, present, and future,” *Color Res. Appl.* **31**, 303–314 (2006).
- A. Gijsenij, T. Gevers, and M. P. Lucassen, “A perceptual comparison of distance measures for color constancy algorithms,” *Computer Vision—ECCV 2008: 10th European Conf. on Computer Vision, Marseille*,

- France, October 12–18, 2008, Proc., Part I 10* (Springer, Cham, 2008), pp. 208–221.
- <sup>11</sup> A. Gijsenij, T. Gevers, and M. P. Lucassen, “Perceptual analysis of distance measures for color constancy algorithms,” *JOSA A* **26**, 2243–2256 (2009).
  - <sup>12</sup> S. Bianco and R. Schettini, “Adaptive color constancy using faces,” *IEEE Trans. Pattern Anal. Machine Intell.* **36**, 1505–1518 (2014).
  - <sup>13</sup> G. Sharma and R. Bala, *Digital Color Imaging Handbook* (CRC Press, Boca Raton, FL, 2017).
  - <sup>14</sup> Y. Hu, B. Wang, and S. Lin, “Fc4: Fully convolutional color constancy with confidence-weighted pooling,” *Proc. IEEE Conf. on Computer Vision and Pattern Recognition* (IEEE, Piscataway, NJ, 2017), pp. 4085–4094.
  - <sup>15</sup> M. Buzzelli, S. Bianco, and R. Schettini, “Arc: Angle-retaining chromaticity diagram for color constancy error analysis,” *JOSA A* **37**, 1721–1730 (2020).
  - <sup>16</sup> M. Buzzelli, “Angle-retaining chromaticity and color space: Invariants and properties,” *J. Imaging* **8**, 232 (2022).
  - <sup>17</sup> P. V. Gehler, C. Rother, A. Blake, T. Minka, and T. Sharp, “Bayesian color constancy revisited,” in *2008 IEEE Conf. on Computer Vision and Pattern Recognition* (IEEE, Piscataway, NJ, 2008), pp. 1–8.
  - <sup>18</sup> G. Hemrit, G. D. Finlayson, A. Gijsenij, P. Gehler, S. Bianco, M. S. Drew, B. Funt, and L. Shi, “Providing a single ground-truth for illuminant estimation for the colorchecker dataset,” *IEEE Trans. Pattern Anal. Machine Intell.* **42**, 1286–1287 (2019).
  - <sup>19</sup> J. Van De Weijer, T. Gevers, and A. Gijsenij, “Edge-based color constancy,” *IEEE Trans. Image Process.* **16**, 2207–2214 (2007).
  - <sup>20</sup> S. Bianco and C. Cusano, “Quasi-unsupervised color constancy,” *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition* (IEEE, Piscataway, NJ, 2019), pp. 12212–12221.
  - <sup>21</sup> D. Cheng, A. Kamel, B. Price, S. Cohen, and M. S. Brown, “Two illuminant estimation and user correction preference,” *Proc. IEEE Conf. on Computer Vision and Pattern Recognition* (IEEE, Piscataway, NJ, 2016), pp. 469–477.