Color Constancy: How to Deal with Camera Bias?

Supplementary Material

Yi-Tun Lin[1,2†]
ethanx188@gmail.com
Bianjiang Yang[3]
yang2208@purdue.edu
Hao Xie[1]
horacezju@gmail.com
Wenbin Wang[1]
wenbinw@meta.com
Honghong Peng[1]
hhpeng@meta.com
Jun Hu[1,4‡]
junhu@apple.com

[1] Meta Platforms, Inc.
CA, USA
[2] Colour and Imaging Lab
University of East Anglia
Norwich, UK
[3] Purdue University
IN, USA
[4] Apple Inc.
CA, USA

A Different Color Correction Methods

In the main paper, we show that camera pre-calibration using homographic color correction [14] can effectively improve the cross-camera generalizability of color constancy algorithms. In fact, there are other effective color correction algorithms commonly discussed in the literature, including linear, polynomial, and root-polynomial least-squares regression (LLS, PLS and RPLS) [16, 29]. Here, we are to examine how these different color correction methods affect the performance of cross-camera color constancy in our evaluation. We test on the simplified C5 model (defined in Section 4 of the main paper), with Canon 5DSR as the color correction reference camera. We use the 2nd-order polynomial and root-polynomial expansions [16] for the PLS and RPLS methods, respectively.

Clearly, in Table A1, we see that homography (homog)—the color correction method used in the main paper—provides leading performance against other color correction methods, on average across all three train/test splits. Regarding the result of each train/test split respectively, we see that all LLS, RPLS and homography pre-calibration can consistently improve the original C5 method. For PLS, it worsens C5's results in the 2nd experiment. We reckon this is because PLS, unlike other three methods, is **not** exposure invariant [16], meaning that the color correction efficacy of PLS can be negatively impacted if the exposure condition under which each image is captured is different from when capturing the training MCC colors. We point readers to [16] for more details on this issue.

[†] The work was done during Yi-Tun's internship with Meta.
[‡] The work was performed when Jun Hu worked at Meta Platforms, Inc.

Method	Mean	Med	Tri	B.25	W.25
(1) Train N+S, Test C					
C5 [8]	2.87	2.24	2.36	0.88	6.00
C5-simp-PLS	2.43	1.57	1.73	0.45	5.92
C5-simp-LLS	2.19	1.46	**1.60**	**0.42**	5.26
C5-simp-RPLS	**2.18**	**1.44**	**1.60**	**0.42**	**5.21**
C5-simp-homog (paper)	2.22	1.46	1.62	**0.42**	5.33
(2) Train C+S, Test N					
C5 [8]	2.68	2.06	2.18	0.59	**5.86**
C5-simp-PLS	2.87	2.09	2.23	0.66	6.45
C5-simp-LLS	2.63	1.61	1.84	0.51	6.52
C5-simp-RPLS	2.59	1.62	1.81	0.51	6.37
C5-simp-homog (paper)	**2.52**	**1.56**	**1.76**	**0.47**	6.23

Method	Mean	Med	Tri	B.25	W.25
(3) Train C+N, Test S					
C5 [8]	3.13	2.50	2.60	0.80	6.62
C5-simp-PLS	2.49	1.66	1.84	0.53	5.90
C5-simp-LLS	2.57	1.64	1.86	0.56	6.16
C5-simp-RPLS	2.61	1.64	1.88	0.60	6.25
C5-simp-homog (paper)	**2.34**	**1.48**	**1.69**	**0.43**	**5.73**
Average of the three train/test splits					
C5 [8]	2.89	2.27	2.38	0.76	6.16
C5-simp-PLS	2.60	1.77	1.93	0.55	6.09
C5-simp-LLS	2.46	1.57	1.77	0.50	5.98
C5-simp-RPLS	2.46	1.57	1.76	0.51	5.94
C5-simp-homog (paper)	**2.36**	**1.50**	**1.69**	**0.44**	**5.76**

Table A1: Changing the color correction methods used for the simplified C5 algorithm. The results of the original C5 algorithm [8] is supplied in the first row of each table for reference. The best results are shown in bold and underlined. C: Canon 5DSR. N: Nikon D810. S: Sony IMX135.

B Training Homography Using a Single Reference Light

In the INTEL-TAU dataset [26], images of the Macbeth Color Checker (MCC) [28] under 10 reference lights captured by the three cameras used are provided. Hence, in the main paper we used all 10 reference illuminations for training (see Section 3.1). Here, we wish to examine how the performance of the homography-corrected color constancy will vary if we only use each 1 out of 10 reference illuminations for the homography training. Again, we evaluate the simplified C5 algorithm with Canon 5DSR as the reference camera.

Method	Mean	Med	Tri	B.25	W.25
(1) Train N+S, Test C					
C5 [8]	2.87	2.24	2.36	0.88	6.00
C5-simp-homog (SL_Hor)	2.55	1.79	1.93	0.57	5.88
C5-simp-homog (IE_A)	2.46	1.69	1.84	0.53	5.73
C5-simp-homog (SL_A)	2.40	1.60	1.76	0.48	5.72
C5-simp-homog (IE_F12)	2.58	1.85	1.98	0.56	5.93
C5-simp-homog (IE_F11)	2.35	1.60	1.74	0.48	5.56
C5-simp-homog (SL_TL84)	2.25	1.43	1.60	0.43	5.50
C5-simp-homog (SL_CW)	2.60	1.89	2.02	0.69	5.77
C5-simp-homog (IE_D50)	2.45	1.60	1.77	0.49	5.90
C5-simp-homog (IE_D65)	2.31	1.50	1.67	0.43	5.60
C5-simp-homog (SL_D)	2.23	**1.42**	**1.59**	0.43	5.40
C5-simp-homog (paper)	**2.22**	1.46	1.62	**0.42**	**5.33**
(2) Train C+S, Test N					
C5 [8]	2.68	2.06	2.18	0.59	**5.86**
C5-simp-homog (SL_Hor)	2.67	1.66	1.88	0.53	6.59
C5-simp-homog (IE_A)	2.89	1.99	2.17	0.60	6.74
C5-simp-homog (SL_A)	2.63	1.66	1.87	0.46	6.50
C5-simp-homog (IE_F12)	2.96	2.12	2.28	0.78	6.63
C5-simp-homog (IE_F11)	2.75	1.81	2.00	0.62	6.51
C5-simp-homog (SL_TL84)	2.57	**1.54**	**1.75**	**0.44**	6.47
C5-simp-homog (SL_CW)	2.67	1.64	1.88	0.49	6.62
C5-simp-homog (IE_D50)	2.74	1.79	2.01	0.70	6.42
C5-simp-homog (IE_D65)	2.61	1.60	1.83	0.48	6.47
C5-simp-homog (SL_D)	2.55	1.60	1.80	0.46	6.32
C5-simp-homog (paper)	**2.52**	1.56	1.76	0.47	6.23

Method	Mean	Med	Tri	B.25	W.25
(3) Train C+N, Test S					
C5 [8]	3.13	2.50	2.60	0.80	6.62
C5-simp-homog (SL_Hor)	2.88	2.30	2.40	0.72	6.13
C5-simp-homog (IE_A)	3.14	2.62	2.70	0.83	6.40
C5-simp-homog (SL_A)	2.73	2.03	2.17	0.59	6.13
C5-simp-homog (IE_F12)	2.66	1.91	2.06	0.57	6.06
C5-simp-homog (IE_F11)	2.51	1.62	1.80	0.54	6.02
C5-simp-homog (SL_TL84)	2.47	1.56	1.75	0.54	6.01
C5-simp-homog (SL_CW)	2.41	1.52	1.71	0.44	5.91
C5-simp-homog (IE_D50)	2.61	1.82	1.98	0.57	6.07
C5-simp-homog (IE_D65)	2.35	**1.47**	**1.65**	0.50	5.74
C5-simp-homog (SL_D)	2.49	1.55	1.74	**0.43**	6.22
C5-simp-homog (paper)	**2.34**	1.48	1.69	**0.43**	**5.73**
Average of the three train/test splits					
C5 [8]	2.89	2.27	2.38	0.76	6.16
C5-simp-homog (SL_Hor)	2.70	1.92	2.07	0.61	6.20
C5-simp-homog (IE_A)	2.83	2.10	2.24	0.65	6.29
C5-simp-homog (SL_A)	2.59	1.76	1.93	0.51	6.12
C5-simp-homog (IE_F12)	2.73	1.96	2.11	0.64	6.21
C5-simp-homog (IE_F11)	2.54	1.68	1.85	0.55	6.03
C5-simp-homog (SL_TL84)	2.43	1.51	1.70	0.47	5.99
C5-simp-homog (SL_CW)	2.56	1.68	1.87	0.54	6.10
C5-simp-homog (IE_D50)	2.60	1.74	1.92	0.59	6.13
C5-simp-homog (IE_D65)	2.42	1.52	1.72	0.47	5.94
C5-simp-homog (SL_D)	2.42	1.52	1.71	**0.44**	5.98
C5-simp-homog (paper)	**2.36**	**1.50**	**1.69**	**0.44**	**5.76**

Table A2: Using individual reference lights provided in the INTEL-TAU dataset [26] to train the homographic corrections used for the simplified C5 algorithm. We order the single-illuminant results based on their CCTs (from low to high). The results of the original C5 algorithm [8] and "C5-simp-homog" in the main paper (where the homographic corrections were trained using all 10 illuminations combined) are supplied in the first and the last row of each table, respectively. The best results are shown in bold and underlined. C: Canon 5DSR. N: Nikon D810. S: Sony IMX135.

Illuminant	CCT (K)
SL_Hor	2330
IE_A	2800
SL_A	2840
IE_F12	2850
IE_F11	3800
SL_TL84	3870
SL_CW	4100
IE_D50	4800
IE_D65	5950
SL_D	6700

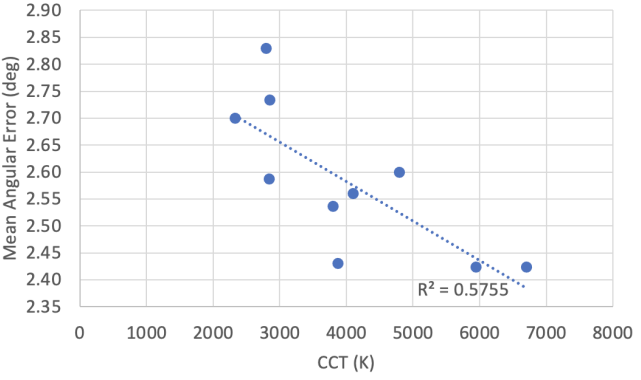

Table A3: (**Left**) The reference illuminants and their corresponding CCTs (Corrected Color Temperatures) provided in the INTEL-TAU dataset [26]. (**Right**) The mean angular error statistics (averaged over the three train/test camera splits) against the CCT of the illuminant used for the homographic correction training.

The numerical results are shown in Table A2. The CCT (Corrected Color Temperature) of each reference illuminant [26], and the plot of averaged mean angular errors against the illuminant CCTs, are provided in Table A3.

Table A2 shows that homographic corrections trained using some illuminations have better effects on solving cross-camera color constancy than others. We observed a general trend that higher CCT illuminants perform better than the lower ones, as shown in the right figure in Table A3. Other factors might also play a role in the performance variation, e.g., the spectral power distributions of the illuminant spectra [26].

Still, on average, we see that simplified C5 adopting homographic corrections with whichever illuminant all performs better than the original C5, though some worst-performing illuminants occasionally get worse than C5 in individual train/test camera split results. Overall, we see that training homographic corrections using all 10 illuminants (as suggested in the main paper) performs the best compared to using single illuminant and the original C5.

C Same-G-Channel Assumption for Edge Colors

In our Same-G-Channel assumption, we assume that the G-channel image of the target camera is going to be the same or a constant scaling apart from the G-channel image for the original camera, for all scenes considered. This is referred to the statistical evidence provided in [20], that most cameras' G-sensor spectral sensitivities can be described by the first PCA basis.

Let us examine the effectiveness of this assumption on predicting the edge color chromaticities of the target camera. Considering the R/G chromaticities in the edge images (and using the Δ symbol to denote the local spatial derivative operation), the Same-G-Channel assumption claims the following approximation:

$$\frac{\Delta R_2}{\Delta G_2} \approx \frac{\Delta(\frac{R_2}{G_2} G_1)}{\Delta G_1}, \tag{1}$$

where the ratio on the left-hand-side is the target edge chromaticity, and on the right is the estimated edge color using the original camera's G-channel value, G_1, and the predicted R_2/G_2 ratio via homography.

It is clear that the approximation can be very accurate if (i) G_1 is a scaling factor apart from G_2, i.e., $G_1 \approx kG_2$ and (ii) the R_2/G_2 predicted by the homographic correction is accurate, as shown in the following derivation:

$$\frac{\Delta(\frac{R_2}{G_2}G_1)}{\Delta G_1} \approx \frac{\Delta\left(\frac{R_2}{G_2}(kG_2)\right)}{\Delta(kG_2)} = \frac{k\Delta R_2}{k\Delta G_2} = \frac{\Delta R_2}{\Delta G_2} \,. \tag{2}$$

Also note that Equation (2) only holds if the local derivative operation (i.e., the Δ) is linear to the constant exposure scaling k. This is true for most works in color constancy, including FFCC [7] and C5 [2].

Next, we wish to examine the effectiveness of the Same-G-Channel assumption in terms of edge color distributions and edge images. For this test, we need same-scene RGB images captured by a set of different cameras. This can be efficiently generated from a set of hyperspectral images and known camera sensitivities [4]. We use images from the ICVL hyperspectral database [3] and camera sensitivities from the RIT 28-camera spectral sensitivity database [20]. And, we train the homographic corrections using 3 reference lights: CIE D65, D50 and A. We visualize some of the best and the worst edge color predictions from Figure A1 to A4.

We observe that the edge chromaticities matching results depend mainly on how well the initial homography mapping performs. In the case of the best results in Figure A1 and A2, the matching is almost identical (by visually comparing the C2 Edge and C1→C2 Edge images). As for the worst results in Figure A3 and A4, we can perceive slight differences especially in the ground-surfaces of the edge images, in which similar color shifts can also be observed when comparing the color images.

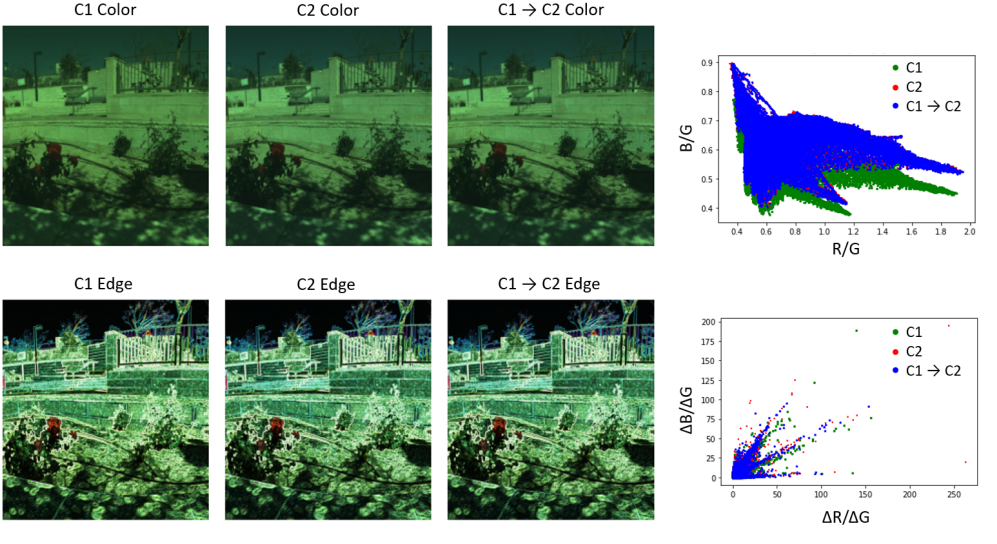

Figure A1: The color (top row) and edge images (bottom row) captured by the original camera C1, target camera C2, and the C2 images predicted from C1 (C1→C2) and with the Same-G-Channel assumption. C1: Pentax K-5, C2: Canon 1DMarkIII. The top panel on the right is the R/G-B/G chromaticity distribution, and the bottom panel is the edge color ($\Delta R/\Delta G$ - $\Delta B/\Delta G$) distribution.

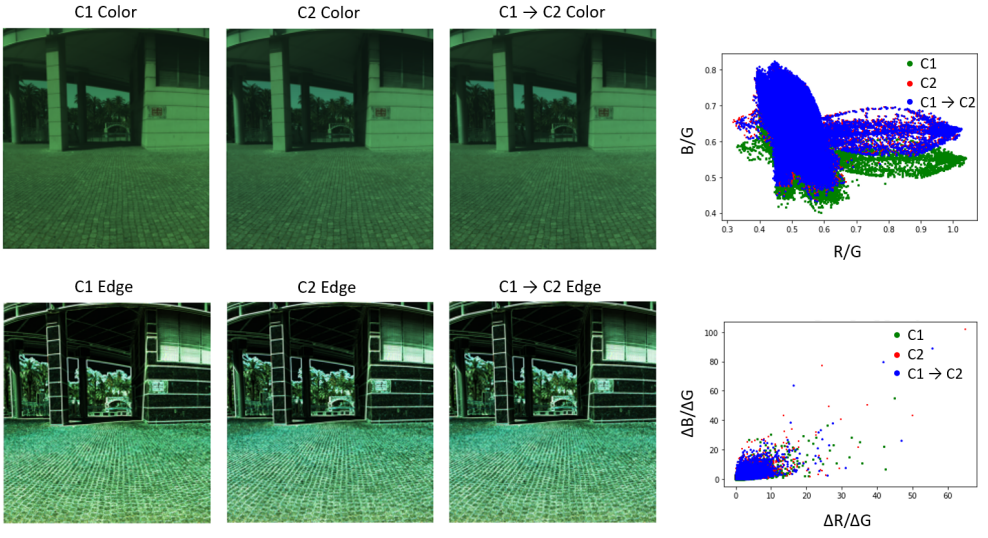

Figure A2: The color (top row) and edge images (bottom row) captured by the original camera C1, target camera C2, and the C2 images predicted from C1 (C1→C2) and with the Same-*G*-Channel assumption. C1: Pentax K-5, C2: Canon 1DMarkIII. The top panel on the right is the R/G-B/G chromaticity distribution, and the bottom panel is the edge color ($\Delta R/\Delta G$ - $\Delta B/\Delta G$) distribution.

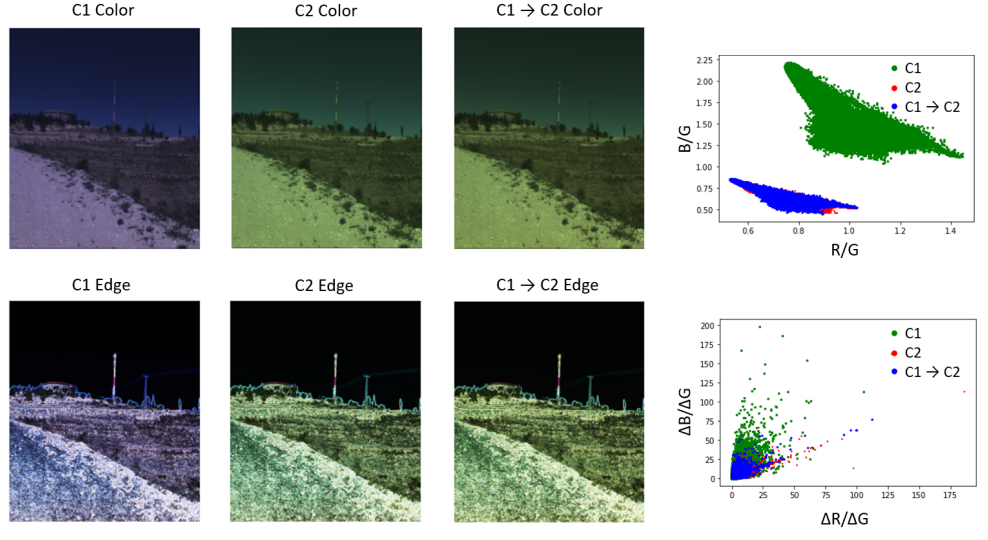

Figure A3: The color (top row) and edge images (bottom row) captured by the original camera C1, target camera C2, and the C2 images predicted from C1 (C1→C2) and with the Same-*G*-Channel assumption. C1: Point Grey Grasshopper 50S5C, C2: Phase One. The top panel on the right is the R/G-B/G chromaticity distribution, and the bottom panel is the edge color ($\Delta R/\Delta G$ - $\Delta B/\Delta G$) distribution.

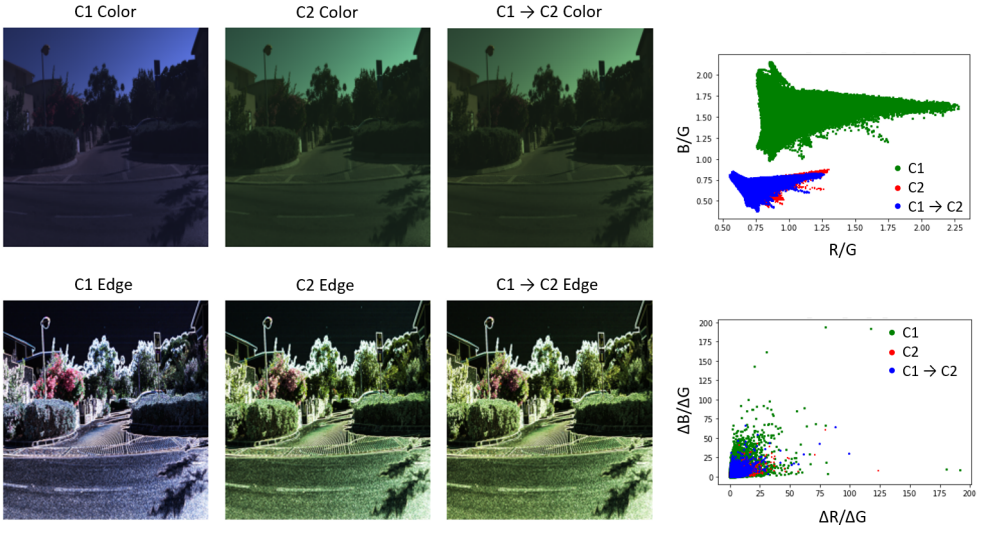

Figure A4: The color (top row) and edge images (bottom row) captured by the original camera C1, target camera C2, and the C2 images predicted from C1 (C1→C2) and with the Same-*G*-Channel assumption. C1: Point Grey Grasshopper 50S5C, C2: Phase One. The top panel on the right is the R/G-B/G chromaticity distribution, and the bottom panel is the edge color ($\Delta R/\Delta G$ - $\Delta B/\Delta G$) distribution.

D Cross-Camera Color Constancy in the Literature

D.1 NUS 8-Camera Database

Leave-the-Testing-Camera-Out Validation. In the literature, there is a collection of validation methodologies used on the NUS database [11, 12] which stresses the idea of training the color constancy algorithms *without* the testing camera's images. However, the actual images used for training vary from one work to another. The examples of this type of validation include the "Leave-One-Out" method in [1, 2] and the "Leave-One-Dataset-Out" method in [22].

Collectively shown in Table A4, we compare the homographic-corrected FFCC with these state-of-the-art methods while separating the methods into categories depending on the training data used. In the first tier we show the results of the statistics-based methods which do not require a training process (and therefore their results are already cross-camera results). In the second tier, we have methods that are trained using three external datasets: Cube+ [5], INTEL-TAU [26] and Gehler-Shi [18], collectively called the data group "A". The third tier includes methods that use not only the images in group A, but also images from 7 out of 8 cameras in the NUS dataset (i.e., only excluding the testing camera) for training. The fourth tier, labeled "Others", consists of only one method [8] which uses *unlabeled* images for training (none of the usual color constancy databases were used). And finally, in the last tier we show the results of the homographic-corrected FFCC (FFCC-homog) and the original FFCC methods we re-ran. Here we use only the 7 out of 8 cameras in the NUS dataset for training. For FFCC-homog, all 5 reference illuminations provided in the NUS dataset are used, and the testing camera in each of the 8-fold experiments was selected as the reference

Category	Method	Mean	Med.	Tri.	B.25%	W.25%
Statistics-based	White Patch [9]	9.91	7.44	8.78	1.44	21.27
	Gray World [10]	4.59	3.46	3.81	1.16	9.85
	Shades-of-Gray [15]	3.67	2.94	3.03	0.99	7.75
	LSRS [17]	3.45	2.51	2.70	0.98	7.32
	2nd-order Gray Edge [31]	3.36	2.70	2.80	0.89	7.14
	1st-order Gray Edge [31]	3.35	2.58	2.76	0.79	7.18
	General Gray World [8]	3.20	2.56	2.68	0.85	6.68
	Gray Pixel (Edge) [32]	3.15	2.20	-	-	-
	Black-and-White PCA [11]	2.93	2.33	2.42	0.78	6.13
	Gray Index [30]	2.91	1.97	2.13	0.56	6.67
A = {Cube+ + INTEL-TAU + Gehler-Shi}	Cross-Dataset FFCC [7]	3.08	2.54	-	-	-
	Cross-Dataset FC4 [7]	2.92	2.34	-	-	-
	FFCC (pixel+edge) [7]	2.87	2.14	2.30	0.71	6.23
	C5 [7]	2.54	1.90	2.02	0.61	5.61
A + NUS	SIIE [4]	2.05	1.50	-	0.52	4.48
	C5 [7]	**1.77**	<u>1.37</u>	<u>1.46</u>	<u>0.48</u>	**3.75**
Others	Quasi-Unsupervised [3]	3.00	2.25	-	-	-
NUS only	FFCC (pixel) [7]	3.10	2.41	2.54	0.84	6.54
	FFCC-homog (pixel)	2.36	1.68	1.82	0.56	5.32
	FFCC (pixel+edge) [7]	2.35	1.74	1.86	0.55	5.19
	FFCC-homog (pixel+edge)	<u>1.80</u>	**1.30**	**1.42**	**0.44**	<u>4.01</u>

Table A4: The NUS's Leave-the-Testing-Camera-Out validation results. The first column indicates the image database(s) used for training (see more details in text). For each statistics, the best result is shown in yellow and bold font, and the second best is shown in cyan and underlined.

camera.

The overall Mean, Median (Med.), Trimean (Tri.), Average of the best 25% (B.25%) and Average of the worst 25% (W.25%) angular error statistics are calculated across the NUS database. The statistics-based results are quoted from Qian et al. [30], and the cross-dataset results are from Afifi et al. [7].

Prior to our proposed FFCC-homog, C5 [7] under the category "A + NUS" performs the best overall. Nevertheless, this is also a category that uses the most amount of data for training. Evidently, compared to C5, FFCC-homog with both pixel and edge information provides on-par performance—superior in the median, trimean and average of the best 25% statistics—while being a much simpler model and using a small fraction of data (we do not amend the group-A images to the NUS) for training.

Cross-Camera Validation. By virtue of the large number of cameras involved in the NUS database, we reckon it is more meaningful to train the FFCC model for each camera at a time and test the model on the other cameras individually.

For each camera used for training, we firstly tune the appropriate training hyperparameters using the usual per-camera 3-fold cross validation setup. Then, applying these hyperparameters, we train an FFCC model on *all* the images from the training camera. Finally, we use this same model to test on the images from other cameras, first without the homographic correction (top of Figure A5) and then with the correction mappings. Here we set the training camera as the reference camera (bottom of Figure A5), and again, all 5 reference illuminations in NUS are used to train the homographic corrections.

For both methods, both pixel and edge information is used in FFCC [7]. Note that in both tables the rows indicate the training cameras while the columns refer to the testing ones. The mean and median angular errors are shown for each pair of training and testing cameras, and

Training Camera		Testing Camera: Can. 1DsMkIII	Can. 600D	Fuj. XM1	Nik. D5200	Oly. EPL6	Pan. GX1	Sam. NX2000	Son. A57
Can. 1DsMkIII	Mean	-	2.2	3.41	3.36	4.06	2.47	3.43	3.54
	Median	-	1.49	3.13	2.75	3.95	1.92	2.86	2.85
Can. 600D	Mean	2.36	-	2.76	2.82	2.92	2.16	2.66	2.9
	Median	1.67	-	2.35	2.24	2.39	1.55	2.06	2.28
Fuj. XM1	Mean	3.22	2.69	-	3.55	2.56	2.89	3.02	4.62
	Median	2.65	2.06	-	3.08	2.13	2.2	2.51	4.42
Nik. D5200	Mean	3.12	2.57	2.63	-	2.65	2.31	2.02	2.26
	Median	2.31	1.86	2.02	-	2.14	1.71	1.45	1.73
Oly. EPL6	Mean	3.59	2.81	2.5	2.85	-	2.69	2.35	3.11
	Median	2.76	2.2	1.79	2.03	-	1.91	1.67	2.03
Pan. GX1	Mean	3.23	2.29	3.7	2.89	3.57	-	2.83	3.28
	Median	2.58	1.61	3.38	2.31	3.6	-	2.32	2.59
Sam. NX2000	Mean	3.12	2.37	2.99	2.2	2.42	2.28	-	2.48
	Median	2.25	1.74	2.21	1.6	1.78	1.65	-	1.85
Son. A57	Mean	4.64	4.33	4.71	2.71	4.0	3.23	2.61	-
	Median	4.28	4.01	4.59	1.88	3.5	2.7	2.16	-

Angular Error: 0 1 2 3 4 5

Training Camera		Testing Camera: Can. 1DsMkIII	Can. 600D	Fuj. XM1	Nik. D5200	Oly. EPL6	Pan. GX1	Sam. NX2000	Son. A57
Can. 1DsMkIII	Mean	-	1.93	2.0	2.03	1.97	2.05	2.13	2.23
	Median	-	1.29	1.34	1.39	1.43	1.62	1.55	1.61
Can. 600D	Mean	2.02	-	1.94	1.76	1.81	2.02	1.91	2.03
	Median	1.45	-	1.4	1.38	1.16	1.53	1.43	1.48
Fuj. XM1	Mean	2.03	1.78	-	1.97	1.75	2.07	2.28	2.44
	Median	1.45	1.25	-	1.33	1.42	1.47	1.82	1.88
Nik. D5200	Mean	2.18	1.6	1.82	-	1.84	2.03	2.01	2.12
	Median	1.45	1.13	1.25	-	1.39	1.56	1.58	1.54
Oly. EPL6	Mean	2.16	1.81	1.82	1.94	-	1.95	1.81	2.12
	Median	1.49	1.32	1.42	1.32	-	1.51	1.37	1.49
Pan. GX1	Mean	2.52	2.43	2.15	2.54	2.46	-	2.94	2.83
	Median	1.88	1.86	1.56	1.95	2.0	-	2.34	2.27
Sam. NX2000	Mean	2.15	1.75	2.12	2.07	1.71	2.08	-	2.0
	Median	1.49	1.18	1.6	1.44	1.03	1.58	-	1.37
Son. A57	Mean	2.41	1.9	2.38	2.18	1.79	2.97	1.92	-
	Median	1.65	1.33	1.76	1.53	1.26	2.52	1.33	-

Angular Error: 0 1 2 3 4 5

Figure A5: The NUS cross-camera validation performance of the original FFCC (**top table**) and FFCC with homographic correction (**bottom table**). Both methods used pixel and edge color information. The shown numbers are angular errors (unit: degree), and the cell color indicates the level of error with respect to the color bar given on the right.

the diagonal cells (where the training and testing cameras are the same one) are left blank.

Clearly, the homographic correction process significantly improves FFCC's performance in this cross-camera testing, or, we say that the correction *stabilizes* the cross-camera performance of FFCC.

D.2 INTEL-TAU 3-Camera Database

Unlike the diverse cross-camera evaluation protocols used on the NUS database in the prior works, upon the proposal of the INTEL-TAU database, it already comes with 2 defined protocols that take the algorithms' cross-camera performance into account [26].

First, the **10-fold cross-validation** protocol (left of Table A5) adopts a non-random separation of the database that aims at lowering the correlation between different folds' images in terms of factors such as the geographical location where the images were taken, illumination conditions, and also, of our concern, the cameras used to capture the images. Then, the **camera invariance** protocol (right of Table A5) is designed with the focus on the cross-camera performance: a 3-fold validation process is set up, with each of the 3 cameras (Canon 5DSR, Nikon D810 and Sony IMX135) in turn used for training, validation and testing[1]. In both experiments, we test FFCC-homog with each of the 3 cameras used as the reference camera. All other evaluation results are quoted from Domislović et al [13].

[1]This evaluation protocol is similar to our proposed one in Section 3.2 in the main paper. However, we reckon that, with data from two cameras at hand for training and validation, one might more naturally choose to mix them up (to increase the data variety in training) and presumably use more data for training than validation. Plus, this protocol does not avoid including same-scene images in both training and testing sets like in our proposed protocol.

		10-fold cross validation					Camera invariance				
Category	**Method**	**Mean**	**Med.**	**Tri.**	**B.25%**	**W.25%**	**Mean**	**Med.**	**Tri.**	**B.25%**	**W.25%**
Statistics-based	Gray World [10]	4.9	3.9	4.1	1.0	10.5	4.7	3.7	4.0	0.9	10.0
	White Patch [8]	9.4	9.1	9.2	1.4	17.6	7.0	5.4	6.2	1.1	14.6
	1st-order Gray Edge [30]	5.9	4.0	4.6	1.0	13.8	5.3	4.1	4.5	1.0	11.7
	2nd-order Gray Edge [30]	6.0	3.9	4.8	1.0	14.0	5.1	3.8	4.2	1.0	11.3
	Shades-of-Gray [9]	5.2	3.8	4.3	0.9	11.9	4.0	2.9	3.2	0.7	9.0
	Black-and-White PCA [20]	4.5	3.2	3.5	0.7	10.6	4.6	3.4	3.7	0.7	10.3
	Weighted Gray Edge	6.1	3.7	4.6	0.8	15.1	6.0	4.2	4.8	0.9	14.2
	Gray Pixel [30]	3.2	2.2	2.4	0.6	7.6	-	-	-	-	-
	Gray Index [30]	3.9	2.3	2.7	0.5	9.8	-	-	-	-	-
Shallow-learning	Color Tiger [8]	4.2	2.6	3.2	1.0	9.9	-	-	-	-	-
	PCC-Q2 [29]	3.9	2.4	2.8	0.6	9.6	-	-	-	-	-
	FFCC [8]	2.4	1.6	1.8	**0.4**?	5.6	-	-	-	-	-
Deep-learning	Quasi-Unsupervised [8]	3.5	2.6	2.8	0.9	7.4	3.4	2.5	2.7	0.8	7.2
	C3AE [29]	3.4	2.7	2.8	0.9	7.0	3.4	2.7	2.8	0.9	7.0
	BoCF [29]	2.4	1.9	2.0	0.7	5.1	2.9	2.4	2.5	0.9	6.1
	FC4 (VGG16) [13]	2.2	1.7	1.8	0.6	4.7	2.6	2.0	2.2	0.7	5.5?
	C5 [8]	2.33	1.55	1.71	0.45	5.57	2.45	1.82	1.95	**0.53**	5.46
	One-net (noise augment) [29]	**1.91**	**1.40**	**1.50**	0.45	**4.25**	**2.1**	**1.6**	**1.7**	**0.5**?	**4.7**
Ours	FFCC-homog (Reference: Canon 5DSR)	2.37	1.58	1.75	**0.43**	5.71	2.59	1.77	1.93	0.58	6.05
	FFCC-homog (Reference: Nikon D810)	2.52	1.65	1.84	0.44	6.08	2.77	1.88	2.08	0.61	6.46
	FFCC-homog (Reference: Sony IMX135)	2.29	1.50	1.69	**0.43**	5.49	2.57	1.73	1.90	0.55	6.06

Table A5: The INTEL-TAU's 10-fold cross validation (**left table**) and camera invariance (**right table**) testing results [26]. For each statistics, the best result is shown in yellow and bold font, and the second best is shown in cyan and underlined. The superscript "?" mark is used for some best results to indicate the lack of precision for determining the detailed ranking.

Evidently, FFCC-homog provides top performances in both benchmarks—second to the One-net method but better than most of the other deep-learning methods. Here, let us discuss on the difference between our method and the One-net method in more details. While in FFCC-homog the cross-camera bias is explicitly corrected using a homography mapping, in contrast, One-net adopted a *noise-augmentation* process which means to apply random Gaussian noises to the ground-truth white points of the original training data—with a standard deviation tuned to the specific cross-camera occasion—which effectively covers the cross-camera shifts of the ground-truth white points [13].

This result is consistent with the main paper's conclusion: single-camera algorithms, such as FC4, CLCC and One-net, can occasionally perform better than cross-camera algorithms like C5 and SIIE (and FFCC-homog where cross-camera bias is pre-corrected). This is likely because of the robust data augmentation applicable to these single-camera methods, as addressed in the One-net publication [13]. This said, with data augmentation, we are at risk of not covering the bias introduced by some corner-case cameras. Advantageously, under the same circumstance, our homographic cross-camera correction method will only require the very-fast retraining of the 3×3 homographic mapping using the MCC chart, without the need for retraining and re-tuning the color constancy backbone.

E Per-Camera Statistics of the Main Paper Results

In main paper's result table (Table 1), we show the averaged results across three train/test camera splits specified in Section 3.2. Here, in Table A6, we show the individual error statistics for each of these train/test splits. Clearly, in accordance to the main paper's results (which are shown in the 4th table), we see that for all testing cameras, all algorithms (except for CLCC in Experiment (2)) are improved using homographic correction. We note that it

is possible that the homographic-corrected data in effect lower the variation in training data which causes the CLCC algorithm to overfit (which can potentially be fixed by adjusting CLCC's regularization hyperparameters).

Method	Mean	Med	Tri	B.25	W.25
(1) Train: N+S, Test: C					
White Patch [0]	11.88	13.70	12.40	1.65	20.73
Shades-of-Gray [2]	6.41	4.37	5.15	1.09	14.84
Gray-World [1]	4.60	3.62	3.90	0.89	9.96
Black-and-White PCA [3]	4.75	3.20	3.54	0.75	11.55
Bright Pixels [4]	4.47	3.11	3.40	0.75	10.51
2nd-order Gray Edge [5]	4.06	3.28	3.45	1.04	8.44
1st-order Gray Edge [6]	4.08	3.21	3.41	0.92	8.77
Gray Index [7]	4.03	2.57	2.89	0.54	10.02
FFCC [7]	4.10	3.57	3.61	0.91	8.43
FFCC-homog	2.30	1.58	1.72	0.51	5.34
FC4 [8]	1.96	1.33	1.45	0.44	4.58
FC4-homog	1.87	1.25	1.38	0.41	4.41
CLCC [9]	**1.72**	1.23	**1.33**	**0.40**	**3.88**
CLCC-homog	1.81	**1.22**	1.34	**0.40**	4.22
SIIE [0]	5.45	4.68	4.87	3.19	8.99
SIIE-homog	2.70	1.86	2.05	0.56	6.30
SIIE-simp-homog	2.75	1.90	2.09	0.59	6.37
C5 [0]	2.87	2.24	2.36	0.88	6.00
C5-homog	2.09	1.39	1.54	0.42	4.96
C5-simp-homog	2.22	1.46	1.62	0.42	5.33
(2) Train: C+S, Test: N					
White Patch [0]	10.06	10.38	9.76	1.61	18.65
Shades-of-Gray [2]	5.43	4.09	4.54	0.99	11.96
Gray-World [1]	5.35	4.22	4.50	1.06	11.59
Black-and-White PCA [3]	4.19	2.88	3.25	0.71	9.92
Bright Pixels [4]	4.26	2.95	3.32	0.72	10.05
2nd-order Gray Edge [5]	4.31	3.47	3.65	1.11	9.00
1st-order Gray Edge [6]	4.26	3.24	3.50	0.98	9.26
Gray Index [7]	3.88	2.08	2.58	0.49	10.19
FFCC [7]	3.17	2.35	2.47	0.63	7.21
FFCC-homog	2.65	1.68	1.91	0.43	6.55
FC4 [8]	2.17	1.46	1.61	0.46	5.13
FC4-homog	2.15	1.49	1.63	0.49	4.98
CLCC [9]	**1.99**	**1.38**	**1.50**	0.44	**4.59**
CLCC-homog	2.05	**1.38**	**1.50**	0.44	4.86
SIIE [0]	4.33	3.91	3.94	1.27	8.20
SIIE-homog	3.23	1.90	2.25	0.53	8.11
SIIE-simp-homog	3.19	1.91	2.25	0.52	8.00
C5 [0]	2.68	2.06	2.18	0.59	5.86
C5-homog	2.34	1.40	1.61	**0.41**	5.89
C5-simp-homog	2.52	1.56	1.76	0.47	6.23

Method	Mean	Med	Tri	B.25	W.25
(3) Train: C+N, Test: S					
White Patch [0]	9.04	9.14	8.90	1.34	16.46
Shades-of-Gray [2]	5.35	4.26	4.55	0.93	11.53
Gray-World [1]	4.72	3.79	4.04	0.97	10.00
Black-and-White PCA [3]	4.58	3.41	3.71	0.74	10.34
Bright Pixels [4]	4.50	3.33	3.64	0.73	10.18
2nd-order Gray Edge [5]	4.19	3.46	3.61	1.01	8.65
1st-order Gray Edge [6]	4.13	3.29	3.48	0.93	8.77
Gray Index [7]	4.09	2.50	2.97	0.53	10.10
FFCC [7]	2.99	2.34	2.43	0.68	6.48
FFCC-homog	2.60	1.67	1.87	0.50	6.34
FC4 [8]	2.43	1.44	1.60	0.45	6.22
FC4-homog	2.20	1.38	1.55	0.43	5.39
CLCC [9]	2.20	1.59	1.69	0.50	5.00
CLCC-homog	**1.96**	**1.29**	**1.42**	**0.39**	**4.69**
SIIE [0]	4.45	3.81	3.94	1.41	8.55
SIIE-homog	3.12	2.09	2.38	0.57	7.32
SIIE-simp-homog	3.10	2.10	2.37	0.54	7.31
C5 [0]	3.13	2.50	2.60	0.80	6.62
C5-homog	2.21	1.41	1.59	0.43	5.36
C5-simp-homog	2.34	1.48	1.69	0.43	5.73
Average of the three train/test splits (main paper results)					
White Patch [0]	10.33	11.07	10.35	1.53	18.61
Shades-of-Gray [2]	5.73	4.24	4.75	1.00	12.78
Gray-World [1]	4.89	3.88	4.15	0.97	10.52
Black-and-White PCA [3]	4.51	3.16	3.50	0.73	10.60
Bright Pixels [4]	4.41	3.13	3.45	0.73	10.25
2nd-order Gray Edge [5]	4.19	3.40	3.57	1.05	8.70
1st-order Gray Edge [6]	4.16	3.25	3.46	0.94	8.93
Gray Index [7]	4.00	2.38	2.81	0.52	10.10
FFCC [7]	3.42	2.75	2.84	0.74	7.37
FFCC-homog	2.52	1.64	1.83	0.48	6.08
FC4 [8]	2.19	1.41	1.55	0.45	5.31
FC4-homog	2.07	1.37	1.52	0.44	4.93
CLCC [9]	1.97	1.40	1.51	0.45	**4.49**
CLCC-homog	**1.94**	**1.30**	**1.42**	**0.41**	4.59
SIIE [0]	4.74	4.13	4.25	1.96	8.58
SIIE-homog	3.02	1.95	2.23	0.55	7.24
SIIE-simp-homog	3.01	1.97	2.24	0.55	7.23
C5 [0]	2.89	2.27	2.38	0.76	6.16
C5-homog	2.21	1.40	1.58	0.42	5.40
C5-simp-homog	2.36	1.50	1.69	0.44	5.76

Table A6: The per-testing-camera cross-camera results on the INTEL-TAU dataset. The homographic-corrected and homographic-corrected-simplified methods are marked in blue and pink, respectively. The best results are shown in bold and underlined.

F Changing Reference Cameras for Homography

In the main paper, all homographic pre-corrections are trained using Canon 5DSR in INTEL-TAU as the reference camera, i.e., all other training/testing camera colors are first transformed to the colors of Canon 5DSR before feeding to color constancy algorithms for training or inference. Here, we compare homographic corrections used on FFCC [7] and C5 [0] when a different reference camera—Nikon D810 or Sony IMX135—is selected.

From the results shown in Table A7, we see that though slight variation exists, both FFCC and C5 are improved using homographic corrections—whichever the reference camera is. We also see that the best-performing reference camera is not fixed: it changes for different train/test camera splits and for different algorithms.

Method	Mean	Med	Tri	B.25	W.25
(1) Train N+S, Test C					
FFCC [7]	4.10	3.57	3.61	0.91	8.43
C5 [2]	2.87	2.24	2.36	0.88	6.00
FFCC-homog (Canon*)	**2.30**	**1.58**	**1.72**	**0.51**	**5.34**
FFCC-homog (Nikon)	2.53	1.79	1.95	0.52	5.77
FFCC-homog (Sony)	2.34	1.65	1.78	0.52	5.43
C5-homog (Canon*)	**2.09**	**1.39**	1.54	0.42	**4.96**
C5-homog (Nikon)	2.13	**1.39**	**1.53**	**0.39**	5.17
C5-homog (Sony)	2.16	1.44	1.59	0.42	5.17
C5-simp-homog (Canon*)	2.22	1.46	1.62	**0.42**	5.33
C5-simp-homog (Nikon)	**2.18**	**1.43**	**1.59**	**0.42**	**5.26**
C5-simp-homog (Sony)	2.22	1.47	1.63	0.44	5.30
(2) Train C+S, Test N					
FFCC [7]	3.17	2.35	2.47	0.63	7.21
C5 [2]	2.68	2.06	2.18	0.59	5.86
FFCC-homog (Canon*)	2.65	1.68	1.91	**0.43**	6.55
FFCC-homog (Nikon)	2.79	1.76	1.99	0.48	6.85
FFCC-homog (Sony)	**2.58**	**1.65**	**1.86**	0.47	**6.33**
C5-homog (Canon*)	2.34	**1.40**	1.61	0.41	5.89
C5-homog (Nikon)	**2.33**	**1.40**	**1.60**	**0.40**	**5.88**
C5-homog (Sony)	2.48	1.55	1.76	0.48	6.11
C5-simp-homog (Canon*)	**2.52**	**1.56**	**1.76**	0.47	**6.23**
C5-simp-homog (Nikon)	**2.52**	**1.56**	**1.76**	**0.46**	6.25
C5-simp-homog (Sony)	2.60	1.62	1.84	0.49	6.42

Method	Mean	Med	Tri	B.25	W.25
(3) Train C+N, Test S					
FFCC [7]	2.99	2.34	2.43	0.68	6.48
C5 [2]	3.13	2.50	2.60	0.80	6.62
FFCC-homog (Canon*)	2.60	1.67	1.87	**0.50**	6.34
FFCC-homog (Nikon)	2.63	1.67	1.87	0.52	6.40
FFCC-homog (Sony)	**2.49**	**1.65**	**1.81**	**0.50**	**5.94**
C5-homog (Canon*)	**2.21**	1.41	**1.59**	0.43	**5.36**
C5-homog (Nikon)	2.25	**1.39**	**1.59**	**0.41**	5.57
C5-homog (Sony)	2.29	1.42	1.64	0.44	5.59
C5-simp-homog (Canon*)	2.34	**1.48**	1.69	**0.43**	5.73
C5-simp-homog (Nikon)	**2.33**	**1.48**	**1.68**	**0.43**	**5.67**
C5-simp-homog (Sony)	2.36	1.49	1.70	0.46	5.74
Average of the three train/test splits					
FFCC [7]	3.42	2.75	2.84	0.74	7.37
C5 [2]	2.89	2.27	2.38	0.76	6.16
FFCC-homog (Canon*)	2.52	**1.64**	1.83	**0.48**	6.08
FFCC-homog (Nikon)	2.65	1.74	1.94	0.51	6.34
FFCC-homog (Sony)	**2.47**	1.65	**1.82**	0.50	**5.90**
C5-homog (Canon*)	**2.21**	1.40	1.58	0.42	**5.40**
C5-homog (Nikon)	2.24	**1.39**	**1.57**	**0.40**	5.54
C5-homog (Sony)	2.31	1.47	1.66	0.45	5.62
C5-simp-homog (Canon*)	2.36	1.50	1.69	**0.44**	5.76
C5-simp-homog (Nikon)	**2.34**	**1.49**	**1.68**	**0.44**	**5.73**
C5-simp-homog (Sony)	2.39	1.53	1.72	0.46	5.82

Table A7: Changing the reference camera used for training the homographic corrections tested on FFCC [7], C5 [2], and simplified C5 algorithms. The results of the original FFCC and C5 algorithms are supplied in the first two rows of each table. For the same color constancy algorithm, the best results among different reference camera used are shown in bold and underlined. The reference camera adopted in the main paper is marked with *. C/Canon: Canon 5DSR. N/Nikon: Nikon D810. S/Sony: Sony IMX135.

References

[1] Mahmoud Afifi and Michael S Brown. Sensor-independent illumination estimation for dnn models. In *British Machine Vision Conference*, 2019.

[2] Mahmoud Afifi, Jonathan T Barron, Chloe LeGendre, Yun-Ta Tsai, and Francois Bleibel. Cross-camera convolutional color constancy. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1981–1990, 2021.

[3] Boaz Arad and Ohad Ben-Shahar. Sparse recovery of hyperspectral signal from natural rgb images. In *European Conference on Computer Vision*, pages 19–34. Springer, 2016.

[4] Boaz Arad, Radu Timofte, Ohad Ben-Shahar, Yi-Tun Lin, and Graham D Finlayson. Ntire 2020 challenge on spectral reconstruction from an rgb image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 446–447, 2020.

[5] Nikola Banić, Karlo Koščević, and Sven Lončarić. Unsupervised learning for color constancy. *arXiv preprint arXiv:1712.00436*, 2017.

[6] Kobus Barnard, Vlad Cardei, and Brian Funt. A comparison of computational color constancy algorithms. i: Methodology and experiments with synthesized data. *IEEE transactions on Image Processing*, 11(9):972–984, 2002.

[7] Jonathan T Barron and Yun-Ta Tsai. Fast fourier color constancy. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 886–894, 2017.

[8] Simone Bianco and Claudio Cusano. Quasi-unsupervised color constancy. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12212–12221, 2019.

[9] David H Brainard and Brian A Wandell. Analysis of the retinex theory of color vision. *Journal of the Optical Society of America A*, 3(10):1651–1661, 1986.

[10] Gershon Buchsbaum. A spatial processor model for object colour perception. *Journal of the Franklin Institute*, 310(1):1–26, 1980.

[11] Dongliang Cheng, Dilip K Prasad, and Michael S Brown. Illuminant estimation for color constancy: why spatial-domain methods work and the role of the color distribution. *Journal of the Optical Society of America A*, 31(5):1049–1058, 2014.

[12] Dongliang Cheng, Brian Price, Scott Cohen, and Michael S Brown. Beyond white: Ground truth colors for color constancy correction. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 298–306, 2015.

[13] Ilija Domislović, Donik Vršnak, Marko Subašić, and Sven Lončarić. One-net: Convolutional color constancy simplified. *Pattern Recognition Letters*, 159:31–37, 2022.

[14] Graham Finlayson, Han Gong, and Robert B Fisher. Color homography: theory and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(1): 20–33, 2017.

[15] Graham D Finlayson and Elisabetta Trezzi. Shades of gray and colour constancy. In *Color and Imaging Conference*, volume 2004, pages 37–41. Society for Imaging Science and Technology, 2004.

[16] Graham D Finlayson, Michal Mackiewicz, and Anya Hurlbert. Color correction using root-polynomial regression. *IEEE Transactions on Image Processing*, 24(5):1460–1470, 2015.

[17] Shaobing Gao, Wangwang Han, Kaifu Yang, Chaoyi Li, and Yongjie Li. Efficient color constancy with local surface reflectance statistics. In *European Conference on Computer Vision*, pages 158–173. Springer, 2014.

[18] Peter Vincent Gehler, Carsten Rother, Andrew Blake, Tom Minka, and Toby Sharp. Bayesian color constancy revisited. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2008.

[19] Yuanming Hu, Baoyuan Wang, and Stephen Lin. Fc4: Fully convolutional color constancy with confidence-weighted pooling. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4085–4094, 2017.

[20] Jun Jiang, Dengyu Liu, Jinwei Gu, and Sabine Süsstrunk. What is the space of spectral sensitivity functions for digital color cameras? In *2013 IEEE Workshop on Applications of Computer Vision*, pages 168–179. IEEE, 2013.

[21] Hamid Reza Vaezi Joze, Mark S Drew, Graham D Finlayson, and Perla Aurora Troncoso Rey. The role of bright pixels in illumination estimation. In *Color and Imaging Conference*, volume 2012, pages 41–46. Citeseer, 2012.

[22] Samu Koskinen, Dan Yang, and Joni-Kristian Kämäräinen. Cross-dataset color constancy revisited using sensor-to-sensor transfer. In *British Machine Vision Conference*, 2020.

[23] Firas Laakom, Jenni Raitoharju, Alexandros Iosifidis, Jarno Nikkanen, and Moncef Gabbouj. Color constancy convolutional autoencoder. In *2019 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 1085–1090. IEEE, 2019.

[24] Firas Laakom, Nikolaos Passalis, Jenni Raitoharju, Jarno Nikkanen, Anastasios Tefas, Alexandros Iosifidis, and Moncef Gabbouj. Bag of color features for color constancy. *IEEE Transactions on Image Processing*, 29:7722–7734, 2020.

[25] Firas Laakom, Jenni Raitoharju, Alexandros Iosifidis, Uygar Tuna, Jarno Nikkanen, and Moncef Gabbouj. Probabilistic color constancy. In *2020 IEEE International Conference on Image Processing (ICIP)*, pages 978–982. IEEE, 2020.

[26] Firas Laakom, Jenni Raitoharju, Jarno Nikkanen, Alexandros Iosifidis, and Moncef Gabbouj. Intel-tau: A color constancy dataset. *IEEE Access*, 9:39560–39567, 2021.

[27] Yi-Chen Lo, Chia-Che Chang, Hsuan-Chao Chiu, Yu-Hao Huang, Chia-Ping Chen, Yu-Lin Chang, and Kevin Jou. Clcc: Contrastive learning for color constancy. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8053–8063, 2021.

[28] Calvin S McCamy, Harold Marcus, James G Davidson, et al. A color-rendition chart. *Journal of Applied Photographic Engineering*, 2(3):95–99, 1976.

[29] Rang Nguyen, Dilip K Prasad, and Michael S Brown. Raw-to-raw: Mapping between image sensor color responses. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3398–3405, 2014.

[30] Yanlin Qian, Joni-Kristian Kamarainen, Jarno Nikkanen, and Jiri Matas. On finding gray pixels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8062–8070, 2019.

[31] Joost Van De Weijer, Theo Gevers, and Arjan Gijsenij. Edge-based color constancy. *IEEE Transactions on Image Processing*, 16(9):2207–2214, 2007.

[32] Kai-Fu Yang, Shao-Bing Gao, and Yong-Jie Li. Efficient illuminant estimation for color constancy using grey pixels. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2254–2263, 2015.