Cross-dataset Color Constancy Revisited Using Sensor-to-Sensor Transfer

Samu Koskinen[1,2]
samu.koskinen@tuni.fi
Dan Yang[2]
dan.yang@tuni.fi
Joni-Kristian Kämäräinen[2]
joni.kamarainen@tuni.fi

[1] Huawei Technologies Oy (Finland) Co. Ltd
Tampere, Finland

[2] Vision Group
Tampere University, Finland

Abstract

Color constancy is required for camera captured images and therefore all digital camera imaging pipelines include an Auto White Balance (AWB) algorithm. An intrinsic problem of AWB is that it is sensor specific and therefore developers need to repeatedly collect new in-house datasets to adjust their methods for new sensors. In literature, the best learning-based methods achieve state-of-the-art performance with clear margin on all available datasets, but performance significantly degrades in cross-dataset experiments due to the aforementioned reason. In this work, we introduce a sensor-to-sensor transfer model that can be used to map datasets with known cameras to any other known camera. The only requirement is that spectral characterizations of the camera models are available. In our experiments, we demonstrate improvements in cross-dataset settings using the proposed sensor-to-sensor transfer model. In addition, for the first time we are able to analyze the characteristics of existing datasets in the common standard observer space and our analysis reveals that certain datasets contain images which are not suitable for color constancy. We introduce a unified cross-dataset color constancy benchmark dataset, compare two state-of-the-art learning-based AWB methods and show superior performance of the proposed sensor-to-sensor model.

1 Introduction

Digital cameras are nowadays used everywhere and for a large number of different scenes and conditions. The color processing pipeline that converts the raw data captured by the camera hardware is tuned separately for each camera model. The different processing blocks of the pipeline are research items on their own, such as demosaic, denoising, tone mapping and color constancy. These research items have been studied for a long time but most of the problems remain unsolved. The color constancy is one of the biggest challenges in the color processing pipe. The learning-based methods have lately become the state-of-the-art solutions for color constancy too [5, 6, 11, 16], but even those algorithms struggle to work in all conditions. One important bottleneck is the amount of training data. Gathering the training data carefully, so that it will cover all the necessary scenes and conditions with accurately measured ground truth information is time consuming and not feasible to do for

each camera model separately. That is one of the main reasons why the learning based color constancy algorithms try to generalize the problem and are not tuned for specific camera models. An intriguing research question is how a camera manufacturer could reuse the existing datasets to train and tune algorithms for a new camera model?

In this work, we propose a sensor-to-sensor image transfer model that can be used to combine and convert existing color constancy datasets into new datasets representing a target camera specified by the user. Camera specific unified datasets enable camera specific training of the learning-based Auto White Balance (AWB) algorithms. This is important from the camera developer point of view as all previously collected datasets can be added to a unified dataset to train better color pipeline algorithms. In addition, we propose illuminant augmentation which further improves the performance of the algorithms.

Contributions

- We propose a Sensor-to-Sensor Transfer model (SST) where images with known illuminant and captured using a known camera are transferred to spectral images and then transferred back to images of another camera.
- We demonstrate the accuracy of the proposed model with images from existing color constancy datasets and by obtaining spectral characterization of camera models employed in the datasets.
- We provide a comparison study of white points in each dataset in the standard observer color space and report several interesting findings.
- Our findings help to select images from the existing datasets to establish a novel *Unified Cross-dataset Color Constancy Benchmark*.
- For two state-of-the-art color constancy methods we obtain significant cross-dataset performance improvements by mapping training images to the target sensor space with SST and by utilizing the SST model for illuminant augmentation.

The Unified Cross-dataset Color Constancy Benchmark will be made publicly available to facilitate fair method comparisons.

2 Related Work

There are only very few works addressing the problem of how to transfer raw images captured with one camera to raw images captured with another camera. We emphasize the prefix *raw* throughout the work since color constancy is our main concern and raw RGB images are heavily biased by the illuminant. The color processing pipelines of digital cameras process raw RGB images to standard RGB images. The standard RGB space (sRGB) is color compensated and therefore colors should appear unbiased as illuminated by the white light (D65). It should be noted that errors produced by Auto White Balance (AWB) algorithms are transferred to the standard RGB space.

An intrinsic part of the proposed model is RGB to spectral image conversion that has been investigated in several works [1, 13, 19, 21, 22, 23]. However, [1, 13, 19, 23] consider only the sRGB to spectral conversion and therefore AWB and other color processing such as the color space transform errors degrade their accuracy. Kawakami *et al.* [21] assume that low resolution hyperspectral images are available in addition to raw RGB images and therefore cannot be used for our purposes. Koskinen *et al.* [22] propose a reverse processing pipeline where large quantities of JPEG images are converted into spectral images, but

their approach assumes that illuminant and camera characteristics are unknown which causes unnecessary metameric problems that our model avoids.

In our work, we assume that spectral characterization of the camera sensor is known. This is a fair assumption from the sensor developer and camera manufacturer point of view. Existing datasets do not provide characterization information, but it can be manually obtained if the camera model(s) are known and can be purchased. There are a number of works that specifically estimate camera spectral response [20, 26, 28], but in our case manual characterization provides more accurate and reliable spectral response information.

The work done by Banić *et al*. [4] describes a method for generating more color constancy test images. The method is capable of generating new test images and also to simulate them as if they were captured with another device. However the method requires manual work, is difficult to automate and depends on the quality of the projected or printed images. Generative Adversarial Networks (GANs) [10] have also been proposed for solving the limited amount of data problem for the color constancy but they have not reached state-of-the-art performance yet. They also generate unpredictable failures and are difficult to debug.

The most related work is done by Nguyen *et al*. [24] who use calibrated linear transforms for conversions between raw images. Their method can work well with certain illuminants but causes clear errors in color space regions that are not well covered by the calibration. Linear matrix conversions are also coarse tools for color space conversions meaning that they cannot do localized color specific transforms that are required for accurate sensor-to-sensor transform. Color constancy databases are sensitive to even small errors which is evident from the several iterations of the popular Shi-Gehler dataset [15].

The proposed Sensor-to-Sensor Transfer model (SST) achieves high accuracy and improves learning-based color constancy in our experiments.

3 Method

3.1 Camera Spectral Characterization

The proposed Sensor-to-Sensor Transfer model (SST) transfers raw RGB images of one camera to raw RGB images of another as if the same scenes would have been captured by two different cameras. The SST requires three type of input data: raw images and their corresponding illuminant (white point) to be transferred, and spectral response characterization of the two cameras. The first two inputs, raw images together with the ground truth

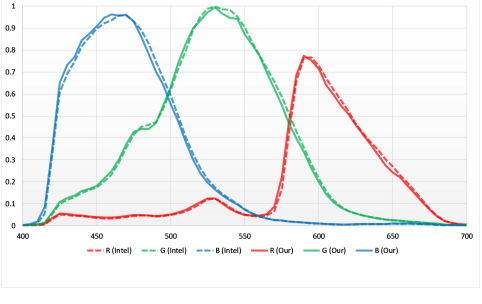

Figure 1: Our vs. authors [2] measurements of spectral responses of a Nikon D810 camera in Intel-TUT dataset.

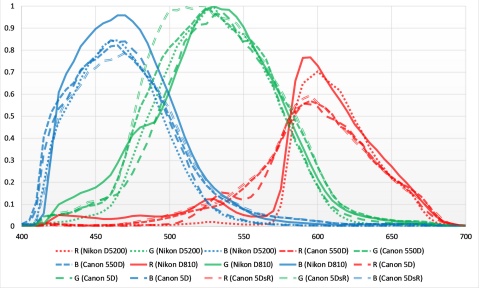

Figure 2: Spectral responses of the cameras selected to our unified cross-dataset benchmark.

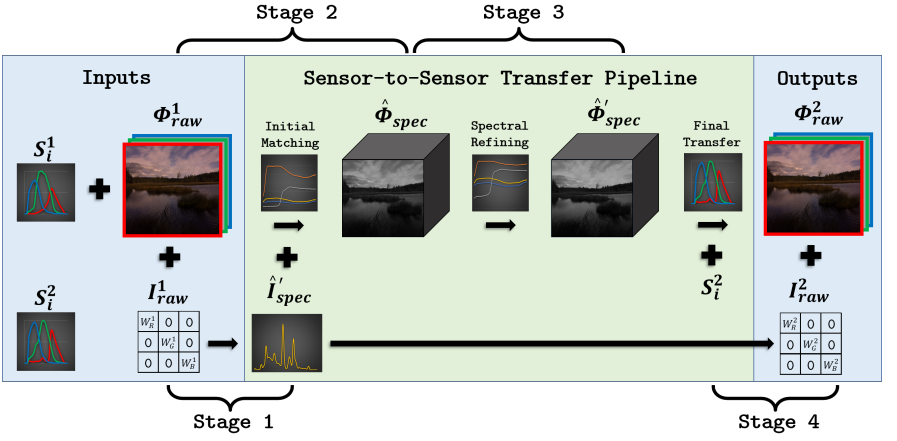

Figure 3: The proposed Sensor-to-Sensor Transfer model (SST) where Φ denotes image, S camera spectral response and I illuminant.

illuminant information, are provided in all commonly used public color constancy datasets. For the experiments, we selected the following four datasets: Shi-Gehler [15], NUS [8], Intel-TUT [2] and Cube+ [3]. However, the third input, spectral camera responses, are more difficult to obtain. Only the Intel-TUT includes the spectral calibration information. However, we can assume that the camera to camera mass production variation of the high-quality DSLR cameras is substantially small. The assumption can be verified by comparing the spectral responses of those provided in Intel-TUT and by measuring another camera of the same model. Our results in Figure 1 verify that we can obtain accurate spectral calibration by measuring the same model even if we do not have exactly the same camera body or know the characteristics of the used lens. On the other hand, the spectral responses differ substantially between different camera models as shown in Figure 2 and thus justifying the SST.

Shi-Gehler contains images captured with Canon 1D and Canon 5D, but in the recent works only the images captured with Canon 5D have been used and that is also the largest part of Shi-Gehler data. We selected the Canon 5D Shi-Gehler images to our experiments as one dataset. NUS dataset contains images captured with 8 different cameras, but since the same scenes are captured with multiple cameras the cross-dataset experiments would be biased. Therefore we selected the Nikon D5200 images from NUS as one dataset. Intel-TUT has the same problem with NUS as each scene was captured with three cameras, two DSLRs and one mobile phone camera. From Intel-TUT we selected the Nikon D810 images. Cube+ images were captured using Canon 550D and thus all Cube+ images were included. We were able to find these camera models and measured their spectral responses. The measurements were made using the Labsphere QES-1000 spectral measurement system. The spectra of the cameras were measured between 400 and 700nm in 5nm steps and they were computed from an average of 50×50 pixel region at the center of each sensor (Figure 2).

3.2 Sensor-to-Sensor Transfer Model

The Sensor-to-Sensor Transfer is an ill-posed problem since the original spectra of the illuminant and scene colors are lost in camera capture using tri-stimulus (RGB) values. Therefore also the SST model provides only estimates of the true raw RGB values. We have care-

fully placed the largest estimation errors at the first processing steps where the RGB data is transformed into the spectral domain. In this transform most of the errors will be seen as metameric errors. Such errors are not of a major concern in the context of color constancy as far as the output raw images are natural and realistic representations of the same scenes including accurate ground truth illuminant white points. Whether some color in the scene is slightly shifted due to metamerism is generally insignificant. Our tests further verify that even the color errors are small. The processing stages of the model are (Figure 3):

1. Illuminant spectrum estimation: I^1_{raw} to $\hat{I}'_{spec}$,
2. Raw to spectral image transform: Φ^1_{raw} to $\hat{\Phi}_{spec}$,
3. Spectral image refinement: $\hat{\Phi}_{spec}$ to $\hat{\Phi}'_{spec}$,
4. Constructions of the final output data: $\hat{\Phi}'_{spec}$ to Φ^2_{raw} and $\hat{I}'_{spec}$ to I^2_{raw} .

Note that for the spectral domain representations we drop the sensor index as the spectral representations $\hat{I}_{spec}$ and $\hat{\Phi}_{spec}$ are canonical and not tied to the sensors 1 or 2. Spectral representations, on the other hand, are only estimates which is denoted by the hat $\hat{}$. The stages are explained in details in the following.

Illuminant spectrum estimation $\hat{I}'_{spec}$ requires a database of real illuminant spectra. They should cover the typical distribution of different light sources. We gathered a set of 100 illuminants containing standard illuminant spectra [18] and some LED spectra we were able to measure. The standard provides an equation for calculating the daylight illuminant spectra with different correlated color temperatures using a weighted combination of existing illuminant spectra: $S(\lambda) = S_0(\lambda) + M_1 S_1(\lambda) + M_2 S_2(\lambda)$ [9]. The same can be done for halogen illuminants by using the Planck's law. This way we were able to generate and collect a good set of 100 illuminants including 70 different daylight spectra spanning 2500-9400K, 13 LED spectra between 2300-5800K, 9 tungsten halogen spectra ranging 2200-3250K and finally 8 fluorescent spectra between 2500-4250K.

The next problems is to pick the correct illuminant spectrum from the illuminant database. This is achieved using the input camera spectral response S^1_i, image illuminant information I^1_{raw} and the illuminant database. We know that the raw image data is formed according to the following equation [29]

$$\Phi_i(x,y) = \int I(\lambda) S_i(\lambda) R(x,y,\lambda) d\lambda, i \in \{\text{R,G,B}\} \quad . \tag{1}$$

As we are only interested to solve the illuminant spectrum in this phase, we set the reflectance spectrum R to a perfect white, i.e. an equal energy spectrum, which in practice means that we can omit both the reflectance and locality (x,y) in (1). We know the camera spectrum S_i and we replace the raw output image Φ_i with the ground truth illuminant I^1_{raw}. Now, all the illuminants I_d from the database are tested and the one that minimizes the equation

$$\hat{I}_{spec} = \arg\min_{I_d} || \int I_d(\lambda) S_i(\lambda) d\lambda - I^1_{raw} ||^2, i \in \{\text{R,G,B}\} \tag{2}$$

is selected. $\hat{I}_{spec}$ is the illuminant spectrum that has a good match with the correct spectral type and color temperature. However, due to the limited size of the illuminant database the match is not perfect but can be fine tuned further. We perform fine tuning using our own empirical method. We apply a virtual hinge to the spectrum with a pivot point at 530nm. From the pivot point the spectrum is linearly adjusted to raise or lower the red and blue parts of the spectrum while still keeping the spectral shape and naturalness intact. Adjustment is

applied iteratively until a perfect match $\hat{I}'_{spec}$ to the ground truth is found according to the equation

$$\hat{I}'^{(t+1)}_{spec}(\lambda) = \hat{I}^{(t)}_{spec}(\lambda)w(\lambda), \tag{3}$$

where w is the hinge weight curve that has a value of 1 at the pivot point. The end points of the vector were matched with the R/G and B/G errors compared to the ground truth illuminant values. The iteration is continued until (2) equals to 0. The optimized illuminant spectrum helps us to define one unknown variable in Equation 1 and thus enables more accurate raw to spectral image transform in the following stages. It is also an essential feature when comparing the white points of different datasets in a common color space (Section 4).

Raw to spectral image transform $\hat{\Phi}_{spec}$ is adapted from the recent work of Koskinen *et al.* [22] by modifying their JPEG to spectrum transform to better suit for the raw RGB. From the previous step a good spectral estimate of the illuminant is obtained and less reverse processing is needed as computation is done in the raw RGB domain. We used the Munsell Glossy database [25] as the source of natural spectra for the optimization.

In this stage we need to get a natural and relatively good match of the spectrum for each pixel. The spectral matching is easier in the CIE L*a*b* color space as the luminance component L* can be omitted and matching done in a 2D space using the Euclidian distance [17]. Our method selects k nearest neighbors and replaces each pixel (x,y) separately with a weighted sum of k Munsell spectra

$$\begin{aligned} \hat{\Phi}_{spec}(x,y) &= \sum_k w_k R^k_{Munsell} \\ \{w_k\} &= \underset{\{w_k\}}{\arg\min} ||\Phi^1_{raw,i}(x,y) - \sum_k w_k \int \hat{I}'_{spec}(\lambda) S^1_i(\lambda) R^k_{Munsell}(\lambda) d\lambda||^2_{a,b} \end{aligned} . \tag{4}$$

Spectral image refinement $\hat{\Phi}_{spec} \rightarrow \hat{\Phi}'_{spec}$ is needed after the initial matching of the spectra. The matching itself does not yet give good enough accuracy for the spectra since it is limited by the gamut and quantity of the spectra in the Munsell database. Therefore the spectra are refined while keeping the shape realistic and natural. The refinement process is conducted in a similar iterative manner as in the illuminant spectrum estimation. However in this case we normalize the camera spectral response curves S^1_i so that the sum of the color channel ($i \in \{\text{R,G,B}\}$) for each wavelength is one. The normalized curves $\bar{S}^1_i$ are used as weights for the refinement process

$$\hat{\Phi}'^{(t+1)}_{spec}(x,y,\lambda) = \hat{\Phi}^{(t)}_{spec}(x,y,\lambda) + \left(\frac{e_i + \varepsilon}{\hat{e}_i} - 1\right)\left(\hat{\Phi}^{(t)}_{spec}(x,y,\lambda) \cdot \bar{S}^1_i(\lambda)\right) , \tag{5}$$

where the channel-wise (RGB) estimation coefficients are $\hat{e}_i$ for the current estimate and e_i for the target. Iteration is continued until $\hat{e}_i = e_i$, i.e. the spectrum matches the raw channel values. ε is a positive constant to make sure that the spectrum is always positive ($\varepsilon = 10^{-6}$ in all our experiments). The target values are directly extracted from Φ^1_{raw} and the estimates are calculated using the Equation 1 by having $I = \hat{I}'_{spec}$, $R = \hat{\Phi}'_{spec}$ and $S = S^1_i$. The stages for converting the raw image into a spectral image enables the SST to support any type of camera sensors without restrictions on their spectral characteristics.

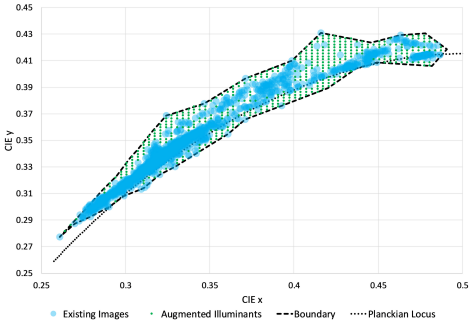

Figure 4: Original (blue) and augmented (green) illuminant white points of NUS

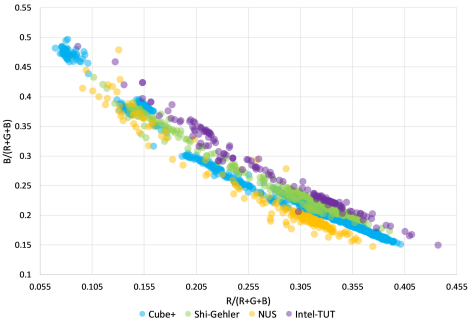

Figure 5: Illuminants of all the used datasets plotted in their own native color spaces

Constructions of the final output data Φ^2_{raw}, I^2_{raw} is straightforward once all the spectra are optimized. Using again Eq. 1, both Φ^2_{raw} and I^2_{raw} are computed with the help of $\hat{I}'_{spec}$, $\hat{\Phi}'_{spec}$ and S^2_i. For the ground truth illuminant computation, we assume a "white" equal energy spectrum as the reflectance spectrum. Therefore that can be omitted from the equation when calculating I^2_{raw}.

3.3 Illuminant Augmentation

It is intriguing to notice that since the spectral representations of scenes and illuminants are canonical and not tied to any sensor, it is easy to add new illuminants to the same scenes. This can be done by simply modifying $\hat{I}'_{spec}$ before constructing the sensor 2 representations in the last step of the model. The most intuitive procedure for illuminant augmentation is to uniformly sample points in the gamut, convex hull, of the original illuminants (white points). The reason to stay within the original boundary is to avoid using unrealistic illuminants or illuminants where color constancy is neither possible nor desired. Augmented white points for the NUS data are illustrated in Figure 4.

4 Cross-dataset Analysis

Our main hypothesis is that learning-based methods for computational color constancy suffer from the fact that sensors in different camera models have different spectral characteristics. The spectral differences of the cameras discussed in Section 3.1 are demonstrated in Figure 2. The spectral differences shift the observed illuminants (white points). Figure 5 shows the dataset white points in the normalized RB-diagram in the datasets' own native color spaces. The diagram verifies that the different spectral responses yield to offsets in the white points. In other words, the same illuminant has different $(\bar{R},\bar{B})$ coordinates in different cameras. It is obvious that this makes cross-dataset experiments challenging for learning-based methods and makes it difficult to analyze what sort of illuminant types are covered by certain regions of the graph.

The *Illuminant spectrum estimation* step in Section 3.2 allows us to draw all illuminants in a common canonical color space such as the CIE xy[14]. In the common space it is easier to analyze the overlap of the illuminants in the different datasets and to identify outliers. The illuminants of all datasets are plotted in Figure 6. There are several interesting findings. The

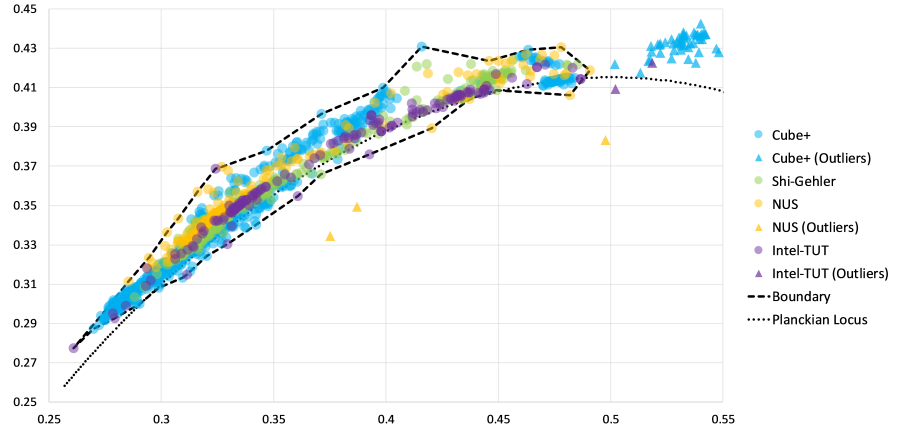

Figure 6: All illuminants transferred to the common CIE xy chromaticity diagram. The boundary of acceptable illuminants and the identified outliers are marked too.

Figure 7: Example of the lab setup test image in the Intel-TUT dataset.

	Angular difference		*Mean color difference*	
Illuminant Type	Orig.	w/ SST	Orig.	w/ SST
D50	6.10	0.23	2.82	1.43
TL84	4.79	2.46	3.81	1.71
CWF	3.80	0.76	3.65	1.19
A	4.79	1.30	3.73	1.09
H	4.34	0.35	3.85	0.79
Mean	4.76	1.02	3.57	1.24
Max	6.10	2.46	3.85	1.71

Table 1: Angular error and color ($\Delta C*$) differences w/ and w/o the proposed SST model.

illuminants are better aligned in Figure 6 than in Figure 5. The datasets seem to contain outlier illuminants that are not well aligned with the Planckian locus. The two NUS outliers at $x < 0.4$ are from restaurant sceneries containing colorful illuminated signs. These scenes thus have multiple white points including some extreme illuminants. In addition, NUS, Cube+ and Intel-TUT contain a number of illuminants at $x > 0.49$ that are captured under sodium vapor street lamps. While this sort of illuminants are commonly used, they are monochromatic inhibiting color vision and making full color constancy impossible. Humans are not able to fully adapt to sodium vapor light [7, 27]. This fact makes sodium vapor illuminated images meaningless for objective color constancy metrics which is why also those were removed from the proposed unified cross-dataset benchmark.

5 Experiments

5.1 Sensor Transfer Accuracy

The first experiment was conducted with the laboratory images from Intel-TUT [7]. The images contain a ColorChecker calibration chart and are illuminated with various standard illuminants generated by the X-Rite Spectralight QC light box as illustrated in Figure 7. We selected two cameras from different manufacturers (spectral responses provided in Intel-TUT): Canon 5DsR and Nikon D810. The Canon images were converted to Nikon raw images using the SST model. The reported performance metrics are the illuminant angular difference

$err = \cos^{-1}(\frac{\underline{\rho}^E \cdot \underline{\rho}^{Est}}{||\underline{\rho}^E|| \cdot ||\underline{\rho}^{Est}||})$ [12] and color difference $\Delta C^* = \sqrt{(a_1^* - a_2^*)^2 + (b_1^* - b_2^*)^2}$ [17]. The angular difference was calculated from the ColorChecker's white patch and color difference using all the color patches.

Table 1 summarizes the results from the first experiment. These verify that SST maps images to another sensor with high accuracy. In fact, even the maximum errors in the angular difference remain below the limit of 3.0 degrees that is largely considered as the threshold of sufficient color constancy. Moreover, the color differences are also very small as the limit of just noticeable difference is 1.0. The small remaining errors are mainly caused by *i*) tri-stimulus metameric errors always present when operating in the spectral domain, *ii*) measurements errors in camera spectral responses, *iii*) measurement inaccuracies of the ground truth illuminant from the raw images and *iv*) errors caused by the iterative optimization routines in SST. However, the results are sufficiently accurate for practical use.

5.2 Cross-dataset Color Constancy

	Unified Test Dataset											
	Cube+			Intel-TUT			NUS			Shi-Gehler		
Method	Mean	Med.	$\leq 3.0°$	Mean	Med.	$\leq 3.0°$	Mean	Med.	$\leq 3.0°$	Mean	Med.	$\leq 3.0°$
FC^4	**2.374**	**1.890**	**73.7%**	3.566	2.974	50.7%	3.105	2.297	62.9%	2.894	2.201	66.2%
FC^4 w/ SST	2.434	1.988	72.0%	3.292	2.672	57.0%	3.354	2.634	59.9%	3.168	2.348	63.7%
FC^4 w/ ill. augm.	2.434	2.092	**73.7%**	**3.024**	**2.343**	64.1%	**2.919**	2.337	**64.0%**	**2.716**	**2.095**	**69.1%**
FC^4 w/ SST+ill. augm.	2.445	2.095	72.4%	3.037	2.373	**66.2%**	3.088	**2.290**	63.5%	3.007	2.282	63.7%
FFCC	2.777	2.440	64.3%	3.303	2.584	58.5%	3.098	2.548	**64.0%**	3.164	3.355	59.8%
FFCC w/ SST	**2.465**	**1.941**	**71.5%**	3.396	2.672	55.6%	3.222	2.513	60.9%	**2.874**	2.139	**66.4%**
FFCC w/ ill. augm.	2.641	2.085	69.4%	**3.122**	2.513	**65.5%**	3.156	**2.242**	62.4%	3.116	2.388	60.2%
FFCC w/ SST+ill. augm.	2.567	2.019	70.1%	3.165	**2.437**	62.7%	**3.080**	2.541	61.4%	2.971	**2.209**	64.7%

Overall results	Avg. Mean	Improvement	Avg. Med.	Improvement	Avg. $\leq 3.0°$	Improvement
FC^4	2.985	-	2.341	-	63.4%	-
FC^4 w/ SST	3.062	-2.52%	2.411	-2.90%	63.2%	-0.36%
FC^4 w/ ill. augm.	**2.773**	**+7.63%**	**2.217**	**+5.58%**	**67.7%**	**+6.86%**
FC^4 w/ SST+ill. augm.	2.894	+3.13%	2.260	+3.56%	66.5%	+4.85%
FFCC	3.086	-	2.732	-	61.7%	-
FFCC w/ SST	2.989	+3.22%	2.316	+17.94%	63.6%	+3.16%
FFCC w/ ill. augm.	3.009	+2.55%	2.307	+18.41%	64.4%	+4.42%
FFCC w/ SST+ill. augm.	**2.946**	**+4.74%**	**2.302**	**+18.69%**	**64.7%**	**+4.99%**

Table 2: Leave-one-dataset-out results of FC^4 [16] and FFCC [5] on the proposed *Unified Cross-dataset Color Constancy Benchmark*. The bottom entries are overall averages.

The Unified Cross-dataset Color Constancy Benchmark was formed by selecting the images of selected cameras in the four publicly available datasets (Section 3.1). The outliers for which color constancy cannot be computed were removed (Section 4). The remaining amount of images in each dataset were 1657 (Cube+), 142 (Intel-TUT), 197 (NUS) and 482 (Shi-Gehler). The two learning-based methods selected to our experiments were FC^4 [16] and FFCC [5], which have reported state-of-the-art results for a number of color constancy datasets. FC^4 is a deep architecture operating on deep semantic features and FFCC is a more conventional method operating on color spaces. The results are summarized in Table 2.

FC^4 benefits from the proposed illumination augmentation in the spectral domain in all four datasets although the differences with Cube+ are only marginal. This finding is supported by the overall results. For example, the number of images for which the estimated white points remain below the threshold 3.0 increases 6.86%. Interestingly, FC^4 operating

on semantic image features does not benefit from the image color transfer. The case is very different with FFCC that operates in the color domain. FFCC benefits from SST with or without illumination augmentation on all four datasets as it cannot automatically cope with the sensor bias. The number of images below the threshold in NUS is lower, but the difference is marginal. Also the FFCC finding is supported by the overall results where SST with illumination augmentation increased the Avg. $\leq 3.0^\circ$ performance by 4.99%. The Cube+, Intel-TUT and Shi-Gehler datasets benefit from SST substantially.

6 Conclusions

Cross-dataset performance of learning-based Auto White Balance (AWB) methods suffer from different sensor characteristics that shift the white points of the same illuminants. We addressed this problem by proposing a Sensor-to-Sensor Transfer model (SST) that can be used to transfer raw RGB images between cameras and to augment more illuminants. For cross-dataset experiments we introduced a novel unified benchmark that was formed from four popular color constancy datasets. SST also allowed us to identify outliers that are unsuitable for color constancy and were thus removed. In the experiments, two state-of-the-art AWB methods, FFCC and FC^4, both benefit from SST and demonstrated substantial improvement on the datasets. Our findings are particularly useful for camera and sensor development as color processing pipeline algorithms for the novel sensors can be adjusted using the previously collected data.

References

[1] B. Arad and O. Ben-Shahar. High-resolution hyperspectral imaging via matrix factorization. In *ECCV*, 2016.

[2] Ç. Aytekin, J. Nikkanen, and M. Gabbouj. INTEL-TUT dataset for camera invariant color constancy research. *CoRR*, abs/1703.09778, 2017.

[3] N. Banić and S. Lončarić. Unsupervised learning for color constancy. *CoRR*, abs/1712.00436, 2017.

[4] N. Banić, K. Koščević, M. Subašić, and S. Lončarić. CroP: Color constancy benchmark dataset generator. *CoRR*, abs/1903.12581, 2019.

[5] J. T. Barron and Y. Tsai. Fast fourier color constancy. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[6] S. Bianco, C. Cusano, and R. Schettini. Color constancy using CNNs. In *CVPR Workshop*, 2015.

[7] M. H. Brill and G. West. Chromatic adaptation and color constancy: A possible dichotomy. *Color Research & Application*, 11(3), 1986.

[8] D. Cheng, D. K. Prasad, and M. S. Brown. Illuminant estimation for color constancy: why spatial-domain methods work and the role of the color distribution. *J. Opt. Soc. Am. A*, 31(5), May 2014.

[9] CIE 015:2018. *Colorimetry*. Commission International de l'Eclairage, 2018.

[10] P. Das, A.S. Baslamisli, Y. Liu, S. Karaoglu, and T. Gevers. Color constancy by GANs: An experimental survey. *CoRR*, abs/1812.03085, 2018.

[11] G. D. Finlayson. Corrected-moment illuminant estimation. In *2013 IEEE International Conference on Computer Vision*, 2013.

[12] G. D. Finlayson, R. Zakizadeh, and A. Gijsenij. The reproduction angular error for evaluating the performance of illuminant estimation algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(7), 2017.

[13] A.S. Glassner. How to derive a spectrum from an RBG triple. *IEEE Computer Graphics and Applications*, 9(4), 1989.

[14] J. Guild and J. E. Petavel. The colorimetric properties of the spectrum. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 230(681-693), 1931.

[15] G. Hemrit, G. Finlayson, A. Gijsenij, P. Gehler, S. Bianco, B. Funt, M. Drew, and L. Shi. Rehabilitating the colorchecker dataset for illuminant estimation. *Color and Imaging Conference*, 2018, 11 2018.

[16] Y. Hu, B. Wang, and S. Lin. FC4: Fully convolutional color constancy with confidence-weighted pooling. In *CVPR*, 2017.

[17] International Organization for Standardization. *ISO 11664-4 Standard - Colorimetry Part 4: CIE 1976 L*a*b* Colour Space*, 2008.

[18] ISO 11664-2:2007(E)/CIE S 014-2/E:2006. CIE colorimetry - part 2: Standard illuminants for colorimetry. Technical report, International Organization for Standardization, 2006.

[19] Y. Jia, Y. Zheng, L. Gu, A. Subpa-Asa, A. Lam, Y. Sato, and I. Sato. From RGB to spectrum for natural scenes via manifold-based mapping. In *ICCV*, 2017.

[20] J. Jiang, D. Liu, J. Gu, and S. Susstrunk. What is the space of spectral sensitivity functions for digital color cameras? In *WACV*, 2013.

[21] R. Kawakami, Y. Matsushita, J. Wright, M. Ben-Ezra, Y.-W. Tai, and K. Ikeuchi. High-resolution hyperspectral imaging via matrix factorization. In *CVPR*, 2011.

[22] S. Koskinen, D. Yang, and J.-K. Kämäräinen. Reverse imaging pipeline for raw RGB image augmentation. In *ICIP*, 2019.

[23] H. Kotera. RGB to spectral image conversion using spectral pallete and compression by SVD. In *ICIP*, 2003.

[24] R. Nguyen, D. K. Prasad, and M. S. Brown. Raw-to-raw: Mapping between image sensor color responses. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014.

[25] J. Orava. The reflectance spectra of 1600 glossy Munsell color chips, 1995. URL https://www3.uef.fi/fi/web/spectral/-spectral-database. Accessed: 2020-04-12.

[26] D. K. Prasad, R. Nguyen, and M. S. Brown. Quick approximation of camera's spectral response from casual lighting. In *2013 IEEE International Conference on Computer Vision Workshops*, 2013.

[27] A. Sharma. *Understanding Color Management*. Graphic Design/Interactive Media Series. Thomson/Delmar Learning, 2004.

[28] P. Urban, M. Desch, K. Happel, and D. Spiehl. Recovering camera sensitivities using target-based reflectances captured under multiple led-illuminations. In *16th Workshop on Color Image Processing*, 2010.

[29] J. von Kries. Influence of adaptation on the effects produced by luminous stimuli. *Source of Color Science*, 1970.