

# Single Pixel Spectral Color Constancy

Samu Koskinen<sup>12</sup>  
samu.koskinen@tuni.fi

Erman Acar<sup>1</sup>  
erman.acar@huawei.com

Joni-Kristian Kämäräinen<sup>12</sup>  
joni.kamarainen@tuni.fi

<sup>1</sup> Huawei Technologies Oy (Finland) Co. Ltd  
Tampere, Finland

<sup>2</sup> Vision Group  
Tampere University, Finland

---

## Abstract

We propose a novel approach for computational color constancy. Instead of raw RGB images used by the existing algorithms to estimate the scene white points, our approach is based on scene average color spectra - a single spectral pixel. We show that as few as 10-14 spectral channels are sufficient. Notably, the sensor output has five orders of magnitude less data than in raw RGB images of a 10MPix camera. The spectral sensor captures the “spectral fingerprints” of different light sources and the illuminant white point can be accurately estimated by a standard regressor. The regressor can be trained with generated measurements using the existing RGB color constancy datasets. To verify the results with real data, we collected a real spectral dataset with a commercial spectrometer. On all datasets the proposed Single Pixel Spectral Color Constancy obtains the highest accuracy in the single dataset and cross-dataset experiments. The method is particularly effective for the difficult scenes for which the average improvements are 40%-70% compared to state-of-the-arts.

## 1 Introduction

A well working color constancy (CC) algorithm is a key component in the camera color processing pipelines. Color constancy is obtained by algorithms that estimate the illuminant white point from captured images. There are static methods that are based on physical or statistical properties of scenes [41, 45] and learning-based methods that learn white point mapping from training data [4, 5, 24]. While the color constancy has been studied for a long time, the problem is not fully solved. Even the best algorithms may fail, for example, when the scene is dominated by a single color.

In this work, we propose a novel approach for computational color constancy. In our approach we replace the raw RGB images used by the existing methods with *average color spectra* of captured scenes. Average spectral sensors are already available in the high-end mobile phones. For example, Huawei P40 Pro is equipped with an 8-channel spectral sensor. It is noteworthy that average spectral measurements completely lack the spatial dimension, but the spectral domain information captures *spectral fingerprints* of illuminants and thus the illuminant white point can be estimated by a simple regression.

The core idea of spectral fingerprints is illustrated in Figure 1. The typical light sources such as a daylight, LED and tungsten are recognizable by the shapes of their power spectra.

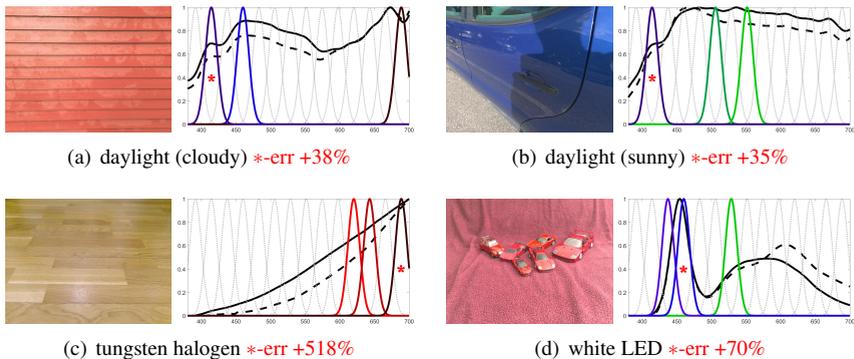


Figure 1: Real spectral dataset examples. Solid black line denotes the light source power spectrum (ground truth) and the dashed is the measured average reflectance spectrum. Gray dotted lines are the 14 spectral channels used in our experiments. For each image the three most important channels found by leave-one-out are colored and the most important denoted by a red asterisk (percentage numbers denote the increase in angular error if this is removed).

The claim can be validated by taking a spectral white point regressor trained with all channels and testing it on unseen images and switching off each channel one by one. The most important channel(s) should be characteristic to each light source. The results for the MLP regressor in Section 3 are shown in Figure 1 for various scenes and light sources. For example, for the both daylight cases the most important channel is the same around 415nm even though the color content of the scenes are very different. That wavelength contains a characteristic bump of the daylight spectrum. For tungsten halogen the important wavelengths are in the near-infrared region. That is characteristic to tungsten sources which have substantial amount of IR energy as compared to their visible light region. The LED case illustrates the practicality of the illuminant fingerprints to identify specific spectral peaks. The blue die peak is captured with the most important channel and the other important channels record more information about the blue peak and the phosphor bump. Other channels are clearly less meaningful in the LED case.

Our main contribution is the novel approach for computational color constancy using scene average color spectrum - a single wide-angle spectral pixel. In addition, we propose **i)** a method to generate spectral data from the existing tristimulus (RGB) color constancy datasets and **ii)** simulation based analysis of optimal spectral sensor design. In all experiments our method obtains lower average angular error than the existing RGB based methods and it is noteworthy that the results are better in cross-dataset experiments where our method is trained with generated data but tested also with real data.

## 2 Related Work

Color constancy algorithms estimate the illuminant  $L$  in order to recover the scene  $R$  under the white light. In the conventional setting  $L$  is estimated from the raw RGB image  $I$ . The existing algorithms can be divided into *learning-free* (static) and *learning-based* methods. In the recent evaluations on multiple datasets [63, 40] the best performing learning-free algorithms are *Gray Index* (GI) [40], *Local Surface Reflectance Statistics* (LSRS) [47], and Cheng *et al.* [40] and the best performing learning-based are *Decoupled Semantic Context and Color Correlation* (DSCCC) [63], *Fast Fourier Color Constancy* (FFCC) [9] and *Fully*

*Convolutional Color Constancy with Confidence* (FC<sup>4</sup>) [24]. The best method varies between the datasets and depending on whether the evaluation is single or cross-dataset, but in overall the differences are small.

There are a few works that study color constancy for (multi)spectral images. For example, Gevers *et al.* [19] use spectral sensing for color constancy assuming that a white reference is available in the scene. Chakrabarti *et al.* [8] model color constancy via spatio-spectral statistics similar to conventional RGB white balance algorithms. Khan *et al.* [54] also extend traditional color constancy algorithms to multispectral images with varying spectral resolutions. These works assume that a full spatial spectral image is available, but compact high resolution spectral cameras are difficult to manufacture. Work done by Chen [9] studies how the Corrected-Moments algorithm [10] can be extended and improved when applied for multispectral images. Spectral sharpening by Finlayson *et al.* [12] aims to improve color constancy with the help of spectral sensing. Hui *et al.* have studied an illuminant source separation task for which they utilize spectral data [25, 26]. Their training data generation in the former paper is physics based and use pre-defined databases for illuminant and reflectance spectra. They also weight their spectral estimation according to a camera spectral response.

Research on spectral measurements is timely as new technological advances make it possible to manufacture miniaturized multispectral sensors. The recent works of Jensen [29] and Wang *et al.* [43] investigate practical implementations of portable spectral sensors.

## 3 Methods

Spectral sensors can be expressed mathematically in a similar way as the RGB sensors of digital cameras. Formation of a raw RGB image  $I$  of a scene  $R$  with the camera  $C$  of known spectral sensitivities  $S_{i=R,G,B}$  and under a global illumination  $L$  can be expressed as [42]

$$I_i(x, y) = \int L(\lambda) S_i(x, y, \lambda) R(x, y, \lambda) d\lambda, i \in \{R, G, B\}, \quad (1)$$

where  $S_i(x, y, \lambda)$  denote the spectral sensitivity of the Red, Green and Blue elements:  $i = \{R, G, B\}$ .  $\lambda$  is the spectral wavelength that for human perceivable colors is 380-700 nanometers (nm). Below 380nm is the ultra-violet band and above 700nm is the infra-red band.

The RGB sensors are designed to capture photographs that match the color sensitive cells of the human visual system (HVS) [69]. However, for accurate color measurements the HVS-inspired wide-band RGB sensors  $C = C^{RGB}$  produce various problems such as the metamerism. The problems can be largely avoided by spectral imaging with a spectral camera  $C^{spec}$  that has multiple narrowband sensor elements  $S_{i=1, \dots, N}$ . Manufacturing of a spectral camera with a high spatial resolution is difficult as it requires a mechanical filter wheel or a large number of photo receptors for each band [16, 57].

### 3.1 Average Spectral Measurement

In this work, we omit the spatial dimension for color constancy. In that case, a spectral camera is not needed. Average spectrum can be measured by a point sensor that needs

- 1) a wide angle lens or diffuser that covers the scene on the image plane  $(x, y)$  of Eq. 1 and
- 2)  $N$  narrowband spectral sensor elements  $S_i$  behind the lens. The sensor  $S_i$  response is

$$\bar{I}_i = \int_x \int_y I_i(x, y) = \int L(\lambda) S_i(\lambda) R(\lambda) d\lambda. \quad (2)$$

The average spectral measurement of a scene  $R$  and under the illumination  $L$  is stored as a vector  $\vec{s} = (\bar{I}_1, \bar{I}_2, \dots, \bar{I}_N)$ . The color constancy problem is to obtain the illuminant  $L$  using

the spectral response vector  $\vec{s}$ . In our simulations,  $\vec{s}$  of only  $N=14$  elements provides good accuracy. This means that sufficient information is available in five orders of magnitude ( $10^5 \times$ ) less data than in a 10MPix camera image.

The field of view (FOV) of the sensor should be as wide as possible in order to integrate and average the changes in the surrounding scenery. This helps to reduce small chromatic objects strongly affecting the shape of the reflected spectrum in a same way as a classic gray world [2] color constancy algorithm works. The field of view should be at least on a same level as the camera's FOV.

## 3.2 Sensor Design

The physical design has restrictions due to the optics, electronics and material properties [2], but for simulation purposes the sensor responses  $S_i$  can be approximated by a Gaussian function,  $Gauss(\mu, \sigma)$ , with the maximum at 1.0 *i.e.* perfect quantum efficiency at the peak wavelength. The Gaussian filter response  $S_i$  is defined by the central wavelength  $\mu_i$  and bandwidth  $\sigma_i$ . The Gaussian spectral shape is a fair assumption also for a practical implementation [29, 43].

Our objective is to find the optimal spectral sensor for color constancy so that it can be implemented in a miniaturized hardware. The number of channels were experimentally tested for  $N= 4, 6, \dots, 16$ . The central bandwidths, Gaussian peaks, were adjusted to uniformly cover the visible spectrum ranging from 380nm to 700nm. This range covers the core of the CIE photopic luminosity function [2]. The channel bandwidth was defined by the *full width at half maximum (FWHM)* and the FWHM bandwidths of 10nm, 20nm and 30nm were tested. These bandwidths were selected to match the capabilities provided by the current technologies. The settings provide 21 different configurations evaluated in Section 5.1.

## 3.3 White Point Regression

The spectral sensor produces a measurement vector  $\vec{s} = (\bar{I}_1, \bar{I}_2, \dots, \bar{I}_N)$  from (2) using the Gaussian responses  $S_i$  (Section 3.2). Color constancy corresponds to an estimation of the global ambient scene illumination  $L = \hat{l} \approx \vec{l}$  [2]. The estimated *white point* is used to normalize the image colors so that achromatic regions appear gray. The white point estimation is defined as a regression problem  $\vec{l} = (l_R, l_G, l_B)^T = f(\vec{s}_{N \times 1})$ , where  $\vec{l}$  is the illuminant white point in RGB and  $f(\cdot)$  is a regression function that maps the spectral measurement  $\vec{s}$  to a white point estimate of  $L$ .

For  $f$  we tested a number of popular regression methods: Kernel Ridge regression (KR) [67], Random Forest regression (RF) [6], and Multilayer Perceptron (MLP) [18]. The Scikit-Learn Python library was used for KR and RF. The methods' hyperparameters were optimized by grid search and cross-validation on the training data and for each sensor configuration separately. MLP was implemented using TensorFlow. MLP has three fully connected hidden layers of sizes 512-1024-512 and the standard Adam optimizer was used. In our experiments the differences between KR, RF and MLP regressors were small and thus any of them is a feasible choice.

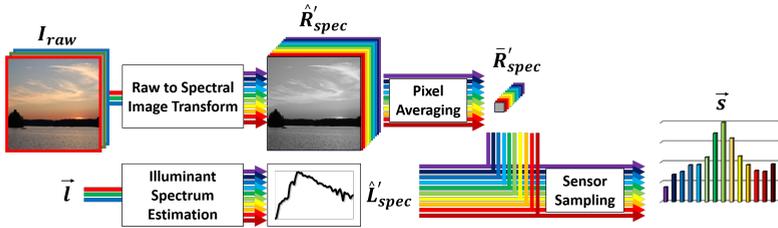


Figure 2: The RGB-to-Spectral conversion model used to generate spectral training data (Section 4.1)

## 4 Data

### 4.1 Generated Spectral Color Constancy Data

In order to train the white point regressors in Section 3.3 we need spectral color constancy training data. It would be straightforward to convert existing spectral image datasets [32, 40, 42] for our purposes, but they are too small and do not contain natural scenes. Alternatively, spectral training data can be generated from the existing color constancy datasets using one of the RGB-to-Spectral conversion methods [10, 30, 51]. The recent Cube+ dataset [8] fits to our purposes. For spectral approximation we adopt parts of the recent Sensor-to-Sensor Transfer (SST) model of Koskinen *et al.* [25]. The original model is designed for RGB-to-RGB conversion between two different RGB sensors and therefore we need to adapt it for RGB-to-Spectral conversion using the following spectral processing steps (Figure 2):

- 1) *illuminant spectrum estimation* ( $\vec{I}$  to  $\hat{L}'_{spec}$ ),
- 2) *raw to spectral image transform* ( $I_{raw}$  to  $\hat{R}'_{spec}$ ),
- 3) *spectral image refinement* ( $\hat{R}'_{spec}$  to  $\bar{R}'_{spec}$ ) and
- 4) *sensor sampling of the average reflected illuminant* ( $\bar{R}'_{spec} \cdot \hat{L}'_{spec}$  to  $\vec{s}$ ).

**Illuminant spectrum estimation.**  $\hat{L}'_{spec}$  estimation is made by finding the closest matching spectrum from an existing database and then refining it to perfectly match the ground truth RGB tristimulus white points in Cube+. For this purpose, we gathered an illuminant database of 100 spectra. Most illuminants were picked from the CIE standard illuminants [27]. The standard does not contain modern LEDs and therefore 13 different LED spectra were measured and added. The standard provides an equation to calculate different daylight spectra as the function of a correlated color temperature:  $L(\lambda) = L_0(\lambda) + M_1L_1(\lambda) + M_2L_2(\lambda)$  [27].  $L_i$  are predefined illuminant characteristics vectors and  $M_i$  are coefficients depending on the selected white point. We selected 70 different daylight illuminants ranging from 2500K to 9400K to cover various conditions from sunsets to cloudy days. The standard also provides typical fluorescent spectra and we selected 8 of those. Finally, we also added 9 tungsten halogen spectra ranging from 2200K to 3250K by using the Planck's law.

As in Eq. 1 the image  $I$  is formed according to [22]:

$$I_i(x, y) = \int L(\lambda) S_i(\lambda) R(x, y, \lambda) d\lambda, i \in \{R, G, B\}. \quad (3)$$

Now that we are only comparing illuminant spectra and the Cube+ ground truth white points, we can set the reflectance spectrum  $R$  to a perfect white and thus effectively omit it from the equation. For the same reason, the spatial information  $(x, y)$  can be removed. We obtained the camera model used in the Cube+ and measured the sensor response spectra  $S_i$  using

Labsphere QES-1000. For spectral matching the image term  $I_i$  is replaced with the ground truth illuminant white point  $\vec{l}$ . Therefore, we only need to find the illuminant  $L_d$  from our database that minimizes the equation

$$\hat{L}_{spec} = \arg \min_{L_d} \left\| \int L_d(\lambda) S_i(\lambda) d\lambda - \vec{l} \right\|^2, i \in \{\mathbf{R}, \mathbf{G}, \mathbf{B}\} . \quad (4)$$

$\hat{L}_{spec}$  is the best match within the 100 illuminants. Since our database contains real illuminant spectra, the best matching illuminant has the natural shape of the corresponding white point. The found spectrum has also similar tristimulus response, but needs fine-tuning. To keep the spectral shape and naturalness intact, refining is done by linearly adjusting the red and blue parts of the spectrum from the pivot point of 530nm. The pivot point is selected to be in the middle of a typical green channel response. The refining is done iteratively until a perfect tristimulus match is achieved for  $\hat{L}'_{spec}$  by utilizing the equation ( $\hat{L}'_{spec} = \hat{L}_{spec}^{(0)}$ )

$$\hat{L}'_{spec}{}^{(t+1)}(\lambda) = \hat{L}'_{spec}{}^{(t)}(\lambda) w(\lambda) , \quad (5)$$

where  $w$  is the weight vector having a value of 1 at 530nm.

**Raw to spectral image transform.** After estimating the illuminant spectrum  $\hat{L}'_{spec} \approx L$ , the only unknown is the scene reflectance spectrum  $R$  in Eq. 3. The same approach from Section 4.1 can be used for reflectance spectrum estimation. The only difference is that the illuminant database is replaced with a database of natural reflectance spectra. The Munsell Glossy dataset [33] is suitable for our purposes. The spectra are well spread over the gamut and the shapes are smooth in nature. Another change we did for the reflectance spectrum estimation is that the matching is made in the CIE  $L^*a^*b^*$  color space [28] where the luminance component  $L^*$  can be omitted. The matching is done in a 2D space using the Euclidean distance. We use  $k$  nearest neighbors and the weighted sum of their Munsell spectra to replace the RGB values of each location  $(x, y)$  with a spectral vector. The results were not very sensitive to selection of  $k$  and thus  $k$  was set to 2 in

$$\begin{aligned} \hat{R}_{spec}(x, y) &= \sum_k w_k R_{Munsell}^k \\ \{w_k\} &= \arg \min_{\{w_k\}} \left\| I_{raw,i}(x, y) - \sum_k w_k \int \hat{L}'_{spec}(\lambda) S_i(\lambda) R_{Munsell}^k(\lambda) d\lambda \right\|_{a,b}^2 . \end{aligned} \quad (6)$$

**Spectral image refinement.** The spectral image refinement is required to perfectly match the Cube+ image RGB values. We normalized the camera spectral responses  $S_i$  so that the sum of the color channels ( $i \in \{\mathbf{R}, \mathbf{G}, \mathbf{B}\}$ ) for each wavelength is one. The normalized curves  $\bar{S}_i$  are utilized as weighting functions for the iteration process ( $\hat{R}'_{spec} = \hat{R}_{spec}^{(0)}$ )

$$\hat{R}'_{spec}{}^{(t+1)}(x, y, \lambda) = \hat{R}'_{spec}{}^{(t)}(x, y, \lambda) + \left( \frac{e_i + \varepsilon}{\hat{e}_i} - 1 \right) \cdot \left( \hat{R}'_{spec}{}^{(t)}(x, y, \lambda) \cdot \bar{S}_i(\lambda) \right) , \quad (7)$$

where the color channel specific (RGB) variables are  $\hat{e}_i$  for the estimate and  $e_i$  for the target. Iteration is finished when the spectrum matches the raw tristimulus values, *i.e.*  $\hat{e}_i = e_i$ . We use  $\varepsilon = 10^{-6}$  to make sure the spectra are always positive. The raw input image  $I_{raw}$  contains the target values and the estimates are calculated using Eq. 3 by placing  $L = \hat{L}'_{spec}$ ,  $S = S_i$  (measured Cube+ camera spectral characterization curves) and  $R = \hat{R}'_{spec}$ .

**Sensor sampling.** In the final step the estimated scene reflectance spectra and the estimated light source spectra are used to construct the spectral sensor response. First, the image spectra are averaged  $\hat{R}'_{spec} \rightarrow \bar{R}'_{spec}$ . The spectral response  $S$  now corresponds to the wide angle multi-channel sensor in Section 3.2 and in the following the index  $i$  refers to the channel number. The final sensor response  $\vec{s}$  is computed from

$$\vec{s} = \int \hat{L}'_{spec}(\lambda) S_i(\lambda) \bar{R}'_{spec}(\lambda) d\lambda . \quad (8)$$

**Noise model.** For more realistic results we added noise to the generated training samples. The noise gives benefit to wider channels with better signal-to-noise levels. The computational spectral sensor channels were defined to have a 100% peak quantum efficiency. We empirically set a very low light condition where the amount of photons to the most sensitive sensor channel is 20 times the FWHM width  $W$  of the channel (in nm). So in effect we assume the same exposure time for each sensor design. We only calculated the photon noise and disregarded the less significant noise sources, such as a read-out noise and ADC noise as those depend heavily on the hardware design which is not known. The photon noise is signal dependent Poisson distributed noise. The strength of the noise can be modeled as a noise which standard deviation grows with a square root of the signal level [15, 21]. Therefore, an equation  $\vec{s} = \vec{s} + \sqrt{20W\vec{s}}X$  was used to add noise to the sensor response  $\vec{s}$  for which most sensitive channel is normalized to one.  $X$  is a random sample from the normal distribution  $\mathcal{N}(\mu, \rho^2) = \mathcal{N}(0, 20W)$ .

**Transform accuracy verification.** In order to verify the accuracy of the used RGB-to-Spectral conversion, we measured the spectral reflectances of the color patches of an X-Rite ColorChecker with a Photo Research PR-670 spectrometer. The spectra were then converted to RGB values using Eq. 1, where the camera spectral sensitivities were from a Huawei Mate 20 Pro and illuminant was set to an illuminant E. The RGB values were then transformed back to spectral values using the proposed RGB-to-Spectral conversion and compared to the original measured ground truth spectra. Any visible errors in the spectral domain are metameric as the differences in the RGB values are negligible. The results are shown in Figure 3 for the challenging saturated content. The average spectra of a scene is typically much less saturated and thus easier for the estimation as indicated by the plotted white patch accuracy.

## 4.2 Real Spectral Color Constancy Data

To validate the results with real data, we collected a spectral color constancy dataset. Each sample contains a raw image captured with a Huawei Mate 20 Pro mobile phone and two spectral measurements by a Konica Minolta CL-70F spectrometer. The first spectral measurement represents the average spectrum of the scene reflected illuminant and the second the ground truth illuminant. The first measurement was made by placing Konica Minolta next to the phone and pointing it towards the scene. The second measurement was made by placing the spectrometer to the scene to measure the ground truth illumination falling on the area. The data gathering setup is illustrated in Figure 5. The ground truth white points were calculated using the illuminant spectrum, the camera spectral response and a perfect white reflectance spectrum in Eq. 3.

The dataset consists of 235 raw images with their corresponding spectral measurements. The dataset was purposely made difficult for color constancy by including scenes that are

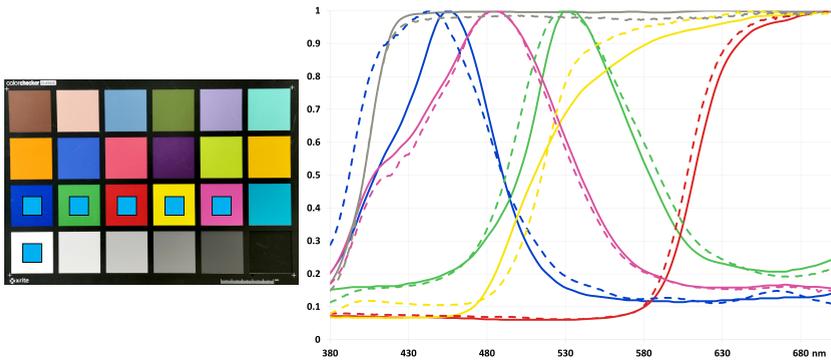


Figure 3: Visualized accuracy of the RGB-to-Spectral conversion. The spectral accuracy is shown for the ColorChecker patches indicated with cyan squares. The ground truths (solid lines) and the estimates (dashed lines) are plotted on the right with colors corresponding to the patches (note that there is no visible difference between the colors of the solid and dashed lines). The spectra were normalized to the peak wavelength.

dominated by a few chromatic colors and often without any clear gray areas. These cases are challenging also to spectral color constancy as the illuminant spectrum and the reflected spectrum are clearly different (the solid and dashed lines in Figure 1). Examples from the dataset are shown in Figures 1 and 6.

## 5 Experiments

### 5.1 Sensor Design

We tested the 21 sensor configurations in Section 3.2: 7 different filter configurations from  $N = 4$  to  $N = 16$  and 3 different filter bandwidths from 10nm to 30nm. The evaluations were made with the generated Cube+ spectral images (Section 4.1) and with the real spectral data (Section 4.2). All results are average numbers from 3-fold cross-validation and the experiments were carried out with noise-free and noise added measurements. The noisy measurements reflect better the performance in realistic low light conditions and demonstrate the difference between the narrow (10nm) and wide (30nm) band sensors. The performance measure in all our experiments is the mean angular error  $err = \cos^{-1} \left( \frac{\vec{l} \cdot \vec{l}'}{\|\vec{l}\| \cdot \|\vec{l}'\|} \right)$  between the ground truth white point  $\vec{l}$  and the estimated white point  $\vec{l}'$  [13].

Results are shown in Figure 4. They provide two expected findings:

- 1) Adding more channels systematically improves the results until they saturate at  $N \geq 10$ .
- 2) Wider filters are more robust to low light and noisy scenes (Cube+).

The average error with the real data ( $\approx 2.4^\circ$ ) is clearly worse than with the generated Cube+ ( $\approx 0.5^\circ$  for clean and  $\approx 1.0^\circ$  for noisy) which can be explained by the fact that the real dataset is smaller and more difficult. However, both results are well below the generally used just noticeable difference of human color perception  $\geq 3.0^\circ$ .

### 5.2 Method Comparison

We compared the spectral color constancy with the settings  $N = 14$  and sensor bandwidth 20nm against three SotA methods: Grayness Index (GI) [41], Fast Fourier Color Constancy (FFCC) [5] and Fully Convolutional with Confidence (FC<sup>4</sup>) [22]. GI is a static method that

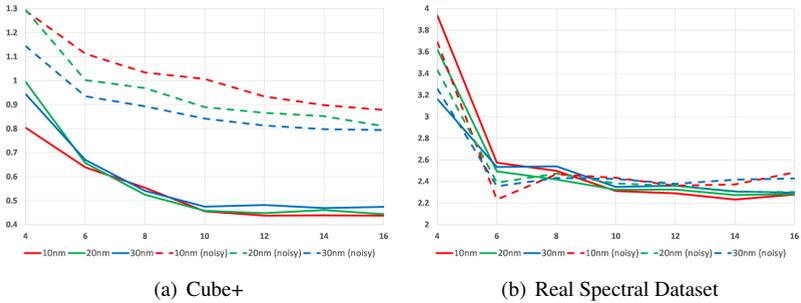


Figure 4: Sensor design results using the MLP white point regressor. Y-axis is the mean angular error from a 3-fold cross-validation and x-axis is the number of channels.

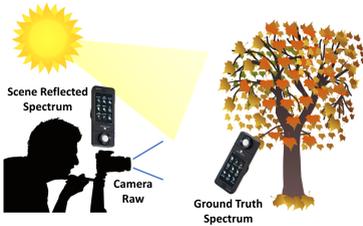


Figure 5: The setup used to capture the real spectral color constancy dataset.

Method	Cube+ [■]			Real Spectral Dataset		
	Mean	Med.	95 <sup>th</sup>	Mean	Med.	95 <sup>th</sup>
GI [■]	2.871	1.664	9.299	7.575	6.512	21.420
FFCC [■]	1.297	0.676	6.420	3.811	2.655	15.288
FC <sup>4</sup> [■]	1.762	1.297	4.608	3.645	2.828	10.282
Spectral (KR)	0.658	0.475	1.640	<b>2.215</b>	<u>1.593</u>	<b>6.244</b>
Spectral (RF)	<u>0.511</u>	<b>0.307</b>	<u>1.541</u>	2.296	1.773	6.567
Spectral (MLP)	<b>0.461</b>	<u>0.312</u>	<b>1.202</b>	<u>2.274</u>	<b>1.584</b>	<u>6.456</u>

Table 1: Comparison of the proposed spectral and SotA color constancy methods in the 3-fold cross-validation. The numbers are angular errors.

does not need training data, but it is competitive against the learning-based methods and particularly effective in cross-dataset evaluations. FFCC and FC<sup>4</sup> are SotA learning-based methods, but with an important difference: FFCC omits the spatial dimension and uses image RGB distributions while FC<sup>4</sup> directly uses the RGB images.

We repeated the 3-fold cross-validation of the previous experiment with the generated Cube+ and the Real Spectral Dataset. The results in Table 1 provide two important findings:

- 1) All three variants of spectral color constancy outperform the SotA RGB methods on both datasets.
- 2) The spectral method is particularly effective on the most difficult scenes (95<sup>th</sup>-percentile) for which it obtains remarkable improvements of 39% to 74%.

Method	NUS [■]			Intel-TUT [■]			Shi-Gehler [■]			Real Spectral Dataset		
	Mean	Med.	95 <sup>th</sup>	Mean	Med.	95 <sup>th</sup>	Mean	Med.	95 <sup>th</sup>	Mean	Med.	95 <sup>th</sup>
GI [■]	2.962	2.103	7.932	3.966	1.923	13.778	3.077	2.168	9.123	7.575	6.512	21.420
FFCC [■]	2.474	1.900	8.457	3.296	2.178	13.044	2.145	1.359	9.048	6.143	4.008	22.333
FC <sup>4</sup> [■]	2.227	1.749	5.581	2.919	2.082	8.416	2.417	1.537	8.331	4.795	<b>3.267</b>	14.597
Spectral (KR)	<u>1.446</u>	<u>0.921</u>	<u>3.344</u>	<b>1.583</b>	<u>1.156</u>	<b>4.213</b>	<u>1.206</u>	0.737	<b>3.966</b>	5.043	4.405	<u>10.239</u>
Spectral (RF)	1.461	1.012	4.179	2.330	1.397	7.663	1.284	<u>0.549</u>	5.614	<b>3.961</b>	<u>3.327</u>	<b>9.056</b>
Spectral (MLP)	<b>1.009</b>	<b>0.792</b>	<b>2.585</b>	<u>1.649</u>	<b>1.044</b>	<u>5.219</u>	<b>1.008</b>	<b>0.491</b>	<u>4.367</u>	<u>4.589</u>	3.937	10.628

Table 2: Angular errors for the cross-dataset experiment. All methods except GI (a static method not requiring training) are trained using the Cube+ images.

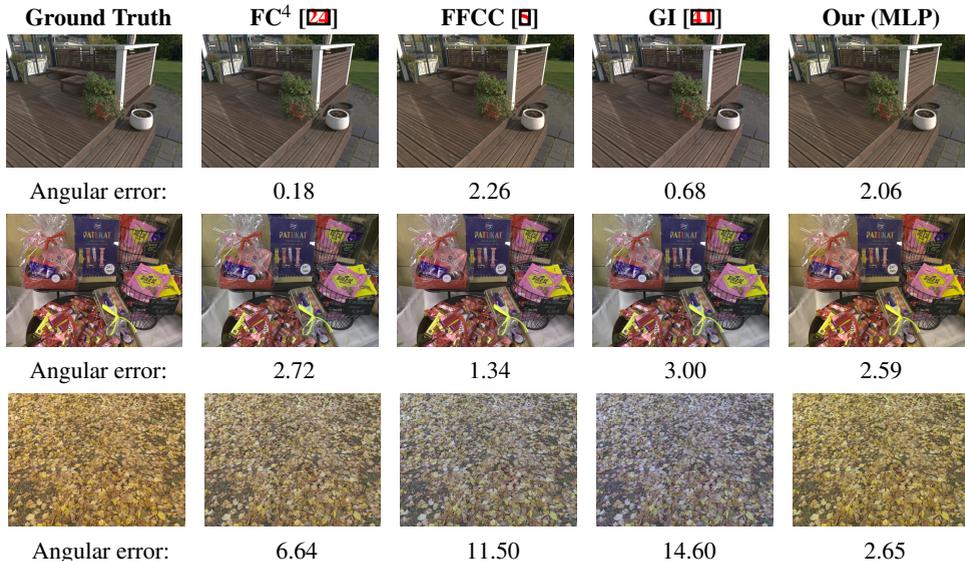


Figure 6: Visualized errors for the tested algorithms. The images include also a static color space transform and an sRGB gamma for displaying purposes.

### 5.3 Cross-dataset Evaluation

The cross-dataset evaluations are important as the methods are not allowed to use training data from the tested datasets and therefore the results reflect better the practical performance. For the cross-dataset evaluations all methods were trained with the Cube+ images. From the popular color constancy benchmarks we selected those where we were able to find the same camera model and measure its spectral response. The selected test datasets were Intel-TUT, NUS and Shi-Gehler, with 142, 197 and 482 images, in addition to our own collected Real Spectral Dataset with 235 images.

The results are shown in Table 2 and visualized in Figure 6. The spectral color constancy method achieved superior or on par accuracy on all four datasets. Similar to the previous experiment, the performance was particularly good for the most difficult images (95<sup>th</sup>-percentile) where the spectral method achieved remarkable improvements between 38%-54%.

## 6 Conclusions

We introduced a new approach for computational color constancy. Instead of the conventional approach using RGB images, our approach uses average color spectra sampled from the visible part of the electromagnetic spectrum. The spectral color constancy achieved the highest accuracy with clear margins to SotA RGB methods. In particular, remarkable improvement of over 50% in the challenging cross-dataset evaluations was achieved with the most difficult cases. It also proved that the data generation method was effective as the results with the generated training data and tested on real measured data still achieved superior results. We conclude that the spectral dimension is more important than the spatial dimension for estimating the illuminant white points.

## 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] J. T. Barron. Convolutional color constancy. In *ICCV*, 2015.
- [5] J. T. Barron and Y. Tsai. Fast fourier color constancy. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [6] L. Breiman. Random forests. *Mach. Learn.*, 45(1):5–32, October 2001. ISSN 0885-6125.
- [7] G. Buchsbaum. A spatial processor model for object colour perception. *Journal of the Franklin Institute*, 310(1):1 – 26, 1980. ISSN 0016-0032.
- [8] A. Chakrabarti, K. Hirakawa, and T. Zickler. Color constancy with spatio-temporal statistics. *IEEE PAMI*, 34(8), 2012.
- [9] X. Chen. *Color Constancy for RGB and Multispectral Images*. Thesis (School of Computing Science). Simon Fraser University, 2017.
- [10] 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.
- [11] G. D. Finlayson. Corrected-moment illuminant estimation. In *2013 IEEE International Conference on Computer Vision*, pages 1904–1911, 2013.
- [12] G. D. Finlayson, M. S. Drew, and B. V. Funt. Spectral sharpening: sensor transformations for improved color constancy. *J. Opt. Soc. Am. A*, 11(5):1553–1563, 1994.
- [13] 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.
- [14] G.D. Finlayson, S.D. Hordley, and P. M. Hubel. Color by correlation: A simple, unifying framework for color constancy. *IEEE PAMI*, 32(11), 2001.
- [15] A. Foi, M. Trimeche, V. Katkovnik, and K. Egiazarian. Practical poissonian-gaussian noise modeling and fitting for single-image raw-data. *IEEE Transactions on Image Processing*, 17(10):1737–1754, 2008.
- [16] L. Gao and L. V. Wang. A review of snapshot multidimensional optical imaging: measuring photon tags in parallel. *Physics reports*, 616(9):1 – 37, 2016.

- [17] S. Gao, W. Han, K. Yang, C. Li, and Y. Li. Efficient color constancy with local surface reflectance statistics. In *Computer Vision – ECCV 2014*, pages 158–173, Cham, 2014. Springer International Publishing. ISBN 978-3-319-10605-2.
- [18] E. H. Geoffrey. Connectionist learning procedures, 1989.
- [19] T. Gevers, H.M.G. Stokman, and J. van de Weijer. Color constancy from hyper-spectral data. In *BMVC*, 2000.
- [20] 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.
- [21] Hamamatsu. *Image Sensors: Selection guide - November 2019*. Hamamatsu Photonics K.K., 2019.
- [22] S. W. Hasinoff. *Photon, Poisson Noise*, pages 608–610. Springer US, Boston, MA, 2014. ISBN 978-0-387-31439-6.
- [23] 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.
- [24] Y. Hu, B. Wang, and S. Lin. FC4: Fully convolutional color constancy with confidence-weighted pooling. In *CVPR*, 2017.
- [25] Z. Hui, K. Sunkavalli, S. Hadap, and A. C. Sankaranarayanan. Illuminant spectra-based source separation using flash photography. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.
- [26] Z. Hui, A. Chakrabarti, K. Sunkavalli, and A. C. Sankaranarayanan. Learning to separate multiple illuminants in a single image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.
- [27] International Organization for Standardization. *ISO 11664-2 Standard - Colorimetry Part 2: Standard Illuminants for Colorimetry*, 2006.
- [28] International Organization for Standardization. *ISO 11664-4 Standard - Colorimetry Part 4: CIE 1976  $L^*a^*b^*$  Colour Space*, 2008.
- [29] K. Jensen. Chip-scale spectral sensing: understanding the new uses for ultra-precise light-source measurement. In *Electronics Know-How*. ams AG, 2020.
- [30] 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.
- [31] 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.
- [32] J. P. Kerekes, K. Strackerjan, and C. Salvaggio. Spectral reflectance and emissivity of man-made surfaces contaminated with environmental effects. *Optical Engineering*, 47(10):1 – 10, 2008.

- [33] V. Keshav and T. P. GVSL. Decoupling semantic context and color correlation with multi-class cross branch regularization. In *2019 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1492–1497, 2019.
- [34] H. A. Khan, J.-B. Thomas, J. Y. Hardeberg, and O. Laligant. Illuminant estimation in multispectral imaging. *J. Opt. Soc. Am. A*, 34(7):1085–1098, 2017.
- [35] S. Koskinen, D. Yang, and J.-K. Kämäräinen. Cross-dataset color constancy revisited using sensor-to-sensor transfer. In *BMVC*, 2020.
- [36] K. Murphy. *Machine Learning: A Probabilistic Perspective*, volume 58. The MIT Press, 01 2012.
- [37] A. H. Nathan and W. K. Michael. Review of snapshot spectral imaging technologies. *Optical Engineering*, 52(9):1 – 23, 2013.
- [38] J. Orava. The reflectance spectra of 1600 glossy Munsell color chips, 1995. URL <https://sites.uef.fi/spectral/databases-software/spectral-database/>. Accessed: 2021-06-19.
- [39] S. E. Palmer. *Vision Science: Photons to phenomenology*. The MIT Press, 1999.
- [40] J. Parkkinen, T. Jaaskelainen, and M. Kuittinen. IEEE 9th international conference on pattern recognition. In *Spectral representation of color images*, volume 2, pages 933–935, 1988.
- [41] Y. Qian, J.-K. Kämäräinen, J. Nikkanen, and J. Matas. On finding gray pixels. In *CVPR*, pages 8054–8062, 2019.
- [42] J. von Kries. Influence of adaptation on the effects produced by luminous stimuli. *Source of Color Science*, 1970.
- [43] Z. Wang, S. Yi, A. Chen, M. Zhou, T. S. Luk, A. James, J. Nogan, W. Ross, G. Joe, A. Shahsafi, K. X. Wang, M. A. Kats, and Z. Yu. Single-shot on-chip spectral sensors based on photonic crystal slabs. *Nature Communications*, 10(1020), 2019.
- [44] S. Westland, A.J. Shaw, and H.C. Owens. Colour statistics of natural and man-made surfaces. *Sensor Review*, 20(1):50–55, 2000.
- [45] K.-F. Yang, S.-B. Gao, and Y.-J. Li. Efficient illuminant estimation for color constancy using grey pixels. In *CVPR*, 2015.