https://doi.org/10.2352/CIC.2025.33.1.19

Enhancing Local Automatic White Balance with Multi-Spectral Imaging

Johannes Keustermans, Robbe Van Beers, Jeroen Hermans, Alex Borgoo and Michael Jacobs; Spectricity (Belgium)

Abstract

Accurate white balancing (WB) remains a critical challenge in image signal processing. Particularly white point estimation for scenes with limited or ambiguous color information can lead to color casts and degraded visual quality. Moreover, many every-day scenes contain multiple relevant illuminants, further complicating illuminant estimation. Estimating multiple white points in a scene exacerbates the challenge and traditional RGB-based WB algorithms struggle with the limited information available for localized regions of a scene. We introduce a novel approach leveraging compact multi-spectral camera technology to improve local WB performance. By capturing additional, narrow-band spectral information beyond the RGB channels, our method provides more accurate white point estimation. We present comparative results demonstrating the advantages of multi-spectral sensing over conventional approaches, highlighting its potential to enable more intelligent and adaptive imaging pipelines in mobile, automotive, and industrial applications. Our approach is based on a relatively simple neural network, trained on simulated multi-spectral measurements. We developed a data collection protocol to establish a medium-sized validation dataset for which we report white point angles and color accuracy (DE2000) values. Our study includes a performance benchmark against a state-of-the art deep learning-based algorithm for our new validation set and the publicly available Large Scale Multi-Illuminant (LSMI) dataset. From the results we observe that our network performs on par with the state-of-the-art algorithm on the LSMI dataset and outperforms the state-of-the-art algorithm on our validation dataset. Furthermore, the multi-spectral based network outperforms the RGB based network.

Introduction

Color constancy is a unique capability of the human visual system to adjust to different lighting conditions to maintain a consistent perception of object colors. It is essential for digital cameras to mimic this capability to result in realistic and visually pleasing images. Therefore, a computational color constancy or white balancing (WB) module is integrated into the imaging pipeline, designed to compensate for the effects of illumination and to recover the true scene colors. Such a white balancing module typically consists of two steps, (1) detecting the white point of the illuminant and (2) replacing this illuminant by a standard illuminant (e.g. D65).

White point detection is a longstanding problem due to its ill-posed nature, thereby requiring prior assumptions about the illuminants and/or scenes. A wide range of methods have been proposed to detect a single white point from an image, ranging from simple heuristics [1], over statistics-based methods [2] to various (deep) learning based methods [3, 4]. In spite of the great progress, the quality of white balancing still highly depends on the captured scene and often large ambiguity remains. Moreover, in reality, a large number of scenes contain multiple illuminants (e.g. indoor scenes with both artificial illumination and daylight, outdoor scenes with direct sunlight and shade, . . .), posing additional challenges to standard white point estimation approaches. Therefore, local white point estimation gained recent interest [5, 6]. While local estimation of illuminant white points has the potential to cope with multi-illuminant scenes, these approaches are confronted with even more ambiguity due to its regional nature.

In this paper, we present a deep learning-based method for local white point detection leveraging multi-spectral image data. Capturing additional narrow-band spectral information beyond the standard RGB channels is expected to reduce the inherent ambiguity of white point detection, especially for scenes with a limited number of colors. The use of multi-spectral information about the scene has been proposed before as an additional input for white point detection [9, 10, 11], however, not for local white point detection.

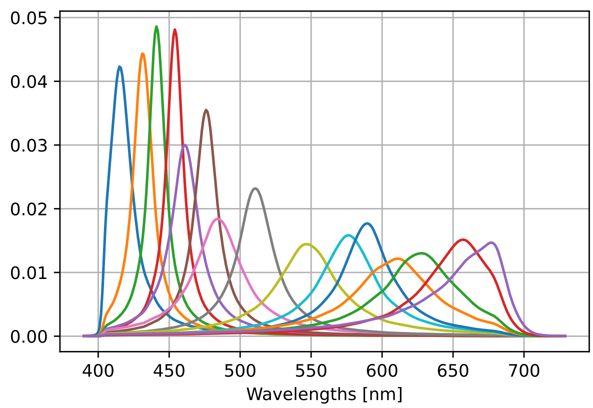

Figure 1. *The spectral sensitivities of Spectricity's S1-M VIS multi-spectral camera module.*

A prerequisite for developing such a method is an extensive and broad dataset of multi-spectral images with accurate local white point information. As to date no such dataset exists, we collected a medium-sized dataset of multi-spectral images using Spectricity's S1-M VIS multi-spectral snapshot camera module. This camera records 15-channel images at SVGA resolution. The spectral sensitivities of this camera module are illustrated in Figure 1. Only sparse local white point information on the measured scenes is present as no in-the-wild applicable protocol exists to measure accurate dense white point maps. This Multi-Illuminant and Multi-spectral Image (MIMI) dataset will be made publicly available[1]. Since multi-spectral measurements from the S1-M VIS camera module can be transformed to accurately mimic the spectral sensitivities of any RGB camera, our dataset can be used to generate ground truth white point data for any RGB camera of interest. The new dataset is used for validation purposes as

[1] `https://spectricity.com/`

training local white point detection methods on such a dataset is susceptible to overfitting and poor generalization, as we will show below. Instead, our method is trained on a much larger dataset of simulated measurements.

We will evaluate a RGB-based version of our local white point detection method on both the publicly available LSMI dataset as well as on our own dataset, and we will benchmark against the LSMI-U network proposed in [5]. We will demonstrate that our simple RGB-based network is competitive with the LSMI-U network, when trained and validated on the LSMI dataset. By evaluating both networks on our own validation dataset, we will illustrate poor generalization capabilities after training on the medium-sized LSMI dataset, and demonstrate improved generalization when trained on our extended simulated dataset.

Finally, we will demonstrate the advantage of multi-spectral imaging for local illuminant detection by comparing a RGB-based version of our method versus a multi-spectral-based one on our MIMI validation dataset.

The remainder of this paper is organized as follows. First, we present the data used for training our local white point detection method, followed by a brief description of the method itself and the validation dataset collected. Finally, results are presented for our local white point detection method along with a comparison to the state-of-the-art.

Methods and data

Training dataset

It is apparent that a performant (local) white balancing method should not be constrained to the types of scenes and illuminants present in the dataset used to train the method and that collecting such a dataset is an arduous task. Therefore, we decided to rely on simulated multi-spectral measurements for model training. More specifically, the network is trained on sets of simulated multi-spectral measurements (network inputs) with corresponding illuminant white points (desired outputs). It is important to note that this set is unordered and no structure is imposed.

Simulated measurements are calculated under the assumption of a perfect Lambertian view of the world where each measurement can be simulated by a combination of an illuminant spectral power distribution (SPD) with a reflectance spectrum projected onto the spectral sensitivities of the camera. Illuminant SPDs are sampled from a large database which includes different artificial illuminant types such as fluorescent, incandescent, halogen and LED lights as well as daylight under different atmospheric conditions and combinations thereof. In total a database of 32768 illuminant SPDs is used. Reflectance spectra are (independently) sampled from an extensive database of 6814 spectra that contains both publicly available as well as in-house measured reflectance spectra.

The training dataset is created as follows. For each set of measurements (typical size is 128) one or more illuminants are selected from the database. In case of multiple illuminants random illuminant mixture coefficients are generated to create an individual illuminant for each measurement. Next, a variable set (between 1 and 15) of reflectance spectra is sampled and randomly assigned to each of the measurements. Subsequently, for each measurement, the respective illuminant SPD and reflectance spectrum are multiplied and projected onto the spectral sensitivities of the respective camera. Finally, the target white points are obtained by projecting the respective illuminant SPDs of each measurement onto the same spectral sensitivities and normalizing the result. Each illuminant (or mixture of illuminants) of the database is used to create 16 different sets of measurements, thus resulting in a training dataset that consists of 524288 sets of measurements. About half of these sets use a single illuminant and half use a mixture of illuminants. After each epoch, the order of the illuminant SPDs is shuffled and new reflectance spectra are sampled. As such, each set of simulated measurements provided for training is different.

While this approach has the potential to generate an abundance of training data, it is clearly a highly simplified model of actual imagery. Indeed, it does not account for any spatial and spectral correlations present in real images, related to both illuminants and reflectance spectra. In addition, it also does not capture more complex light-matter interactions beyond simple Lambertian reflection. Despite these drawbacks of our training data, we will demonstrate that we obtain competitive local illuminant estimation results on real-world data.

Network

Most existing local white point detection methods are based on large deep learning networks which exceed the computational capabilities of mobile devices [5, 6, 7]. Therefore, our method is based on a relatively simple transformer-based neural network with a limited number of parameters. It consists of four blocks: i) an encoder that processes each measurement individually, ii) an aggregation block that extracts the meaningful illuminant information, iii) a dispersion block that relates the collected illuminant information to each of the inputs and iv) a decoder that converts the output of the dispersion block into white points. See Figure 2 for an illustration. The encoder and decoder are simple feed forward networks. The aggregation and dispersion blocks use the set transformer architecture [12].

Networks are trained for 100 epochs with an initial learning rate of $1.0e^{-4}$ that is halved when the validation loss does not sufficiently decrease after 5 epochs. Data augmentation during training includes small random deformations of the illuminant SPD and adding realistic shot noise to the simulated measurements.

This method is not restricted to multi-spectral measurements. It is equally possible to train it on RGB measurements. To illustrate the potential of our method we use two variants: (1) start from a set of RGB measurements and predict the corresponding RGB white points for a given camera, (2) start from a set of multi-spectral measurements and predict the corresponding multi-spectral white points for our camera. Also it is straightforward to fine-tune our method on real measurement data.

Validation dataset

Ideally, local white balancing algorithms are trained and tested on a dataset that contains the ground truth white point at every pixel in the image. The LSMI dataset [5] was established to meet this requirement for RGB data. We found that the presented dense white point maps are not always sufficiently accurate, partly due to the approximations made in the employed imaging model. In addition, the requirement to control all-but-one illuminant is not possible in many realistic conditions (e.g. see Figure 3). Moreover, when used for training local white point estimation methods, we show in the results below that the limited number of data samples and poor splitting between training and test set leads to overfitting.

To illustrate the underlying approximations of the LSMI imaging model, consider the following expressions below. In these expression **m** is a measurement, **S** is the matrix with the

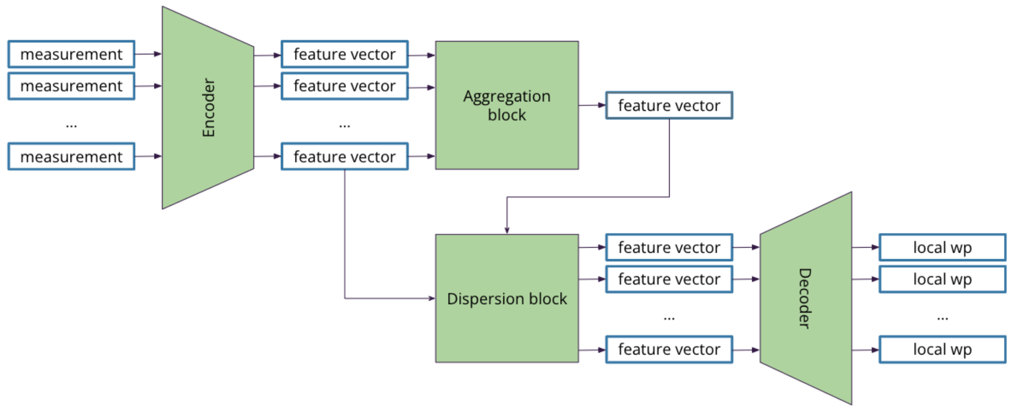

Figure 2. High level overview of our network architecture which starts with a feed forward encoder block, followed by a set transformer based aggregation and dispersion block and ends with a feed forward decoder block.

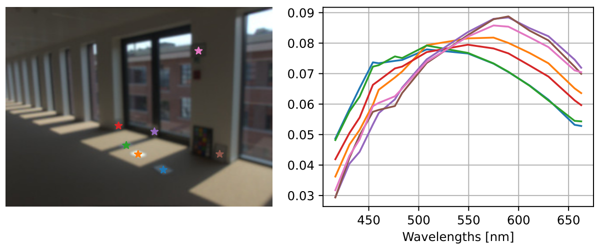

Figure 3. Left: exemplar scene where not all-but-one illuminants can be controlled. In this scene the LED illumination on the ceiling can be controlled but the daylight entering through the window and the cast shadow cannot be controlled. Right: the (multi-spectral) white points measured at the white charts in the scene (colors correspond). Please note the difference between white charts in direct sunlight and white charts in the shade.

spectral sensitivities of the camera, $\Lambda_r = \mathrm{diag}(\mathbf{r})$ is a diagonal matrix with the reflectance spectrum $\mathbf{r}$ on its diagonal and λ is the spectral power distribution of the illuminant

$$
\begin{aligned}
\mathbf{m} &= \mathbf{S}\Lambda_r\lambda \ , && (1)\\
&\approx \mathbf{S}\Lambda_r\mathbf{S}^{+}\mathbf{S}\lambda \ , &&\\
&= \mathbf{D}_r\mathbf{l} \ , &&\\
&\approx \mathbf{D}_r\mathbf{1}\odot\mathbf{l} \ , &&\\
&= \mathbf{d}_r\odot\mathbf{l} \ , && (2)
\end{aligned}
$$

where $\mathbf{S}^{+}$ is the pseudo-inverse of $\mathbf{S}$, $\mathbf{D}_r = \mathbf{S}\Lambda_r\mathbf{S}^{+}$, $\mathbf{l} = \mathbf{S}\lambda$ is the white point of the illuminant and $\mathbf{d}_r = \mathbf{D}_r\mathbf{1}$. Expression 2 is essentially the imaging model employed by the LSMI protocol (expression 1 from [5]). These approximations are only valid if the illuminant SPD fully lives in the subspace defined by $\mathbf{S}$ (in this case $\mathbf{S}^{+}\mathbf{S}\lambda = \lambda$) and if $\mathbf{D}_r$ is a diagonal matrix. The former is not likely to be true, especially for a RGB camera. Please note that this approximation is much more accurate for a multi-spectral camera.

Although we train our methods on simulated measurements, a validation dataset consisting of real (multi-spectral) images is needed for validation. The goal is to establish an annotated dataset that can be used to compare the performance of our multi-spectral local white balancing algorithm with a similar RGB-based algorithm and state-of-the-art RGB-based algorithm. Here we chose to determine a sparse set of accurate ground truth white points, rather than the approximate dense white point maps of [5]. This is both sufficient for our purpose and straightforward to collect reliably.

The ground truth white points are determined on white charts placed in the scene. In this manner we can perform the local white point estimation on a first image without white charts and then compute errors (white point angles) with respect to the measurements on carefully placed white charts in a second image. Charts were placed in front of a flat surface to make the measurement of the white point on that surface as accurate as possible. The final dataset contains 774 scenes with pairs of images for each device. Each image pair consists of an image with and an image without white charts. See Table 1 for further details of the dataset content. A Macbeth color chart was also added to the second image to quantify color accuracy after white balancing. In order to gain an idea of the broad performance of our white point detection algorithms, an effort was made to collect a large variety of scenes: inside, outside, office, home, park, etc. A data sample for one scene is shown in Figure 4.

Overview of the data available in the MIMI (validation) dataset.

	Samsung S9 RGB camera	Spectricity S1-M VIS camera
Recorded scenes	774	774
Data	Raw, jpeg	Raw, cube
Android Camera2 CaptureResults	✓	
Annotated white charts	✓	✓
Annotated Macbeth chart	✓	✓
Device calibration data	✓	✓

We mounted 2 Samsung Galaxy Tab S9 tablets with Spectricity S1-A accessory devices on a tripod as shown in Figure 5. Stability of the setup is crucial for reliable comparison of consecutive images with and without white charts. For each scene recording proceeded as follows. First the tripod with mounted devices was pointed at a scene and data capture was triggered by a remote to avoid accidental movement of the setup. Next the charts were placed in front of surfaces in the scene. Where possible flat surfaces were preferred. Finally, a second capture was triggered. After data collection the images were manually annotated to indicate the position of each white chart and the Macbeth color chart. See Figure 4 for the resulting annotations, (sparse)

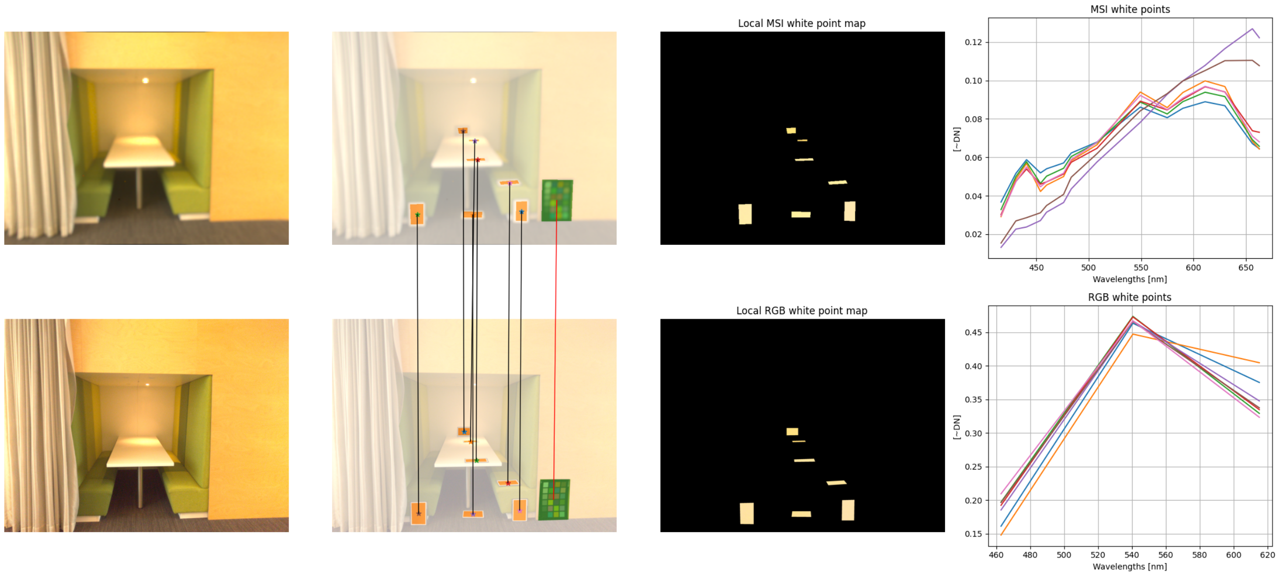

Figure 4. *Top row is the result for the S1-M VIS (multi-spectral) camera module, bottom row for the RGB camera from the Samsung Galaxy Tab S9 tablet. From left to right: non-white balanced sRGB image of the scene without white charts, non-white balanced sRGB image of the scene with white charts and annotations, local white point maps in sRGB space and finally, local white points.*

local white point maps and white points.

Results

RGB-based white point detection

Since the LSMI dataset and associated local white point detection network (LSMI-U) [5] are recognized as state-of-the-art and publicly available, we evaluate our method on this dataset and benchmark against the LSMI-U network. This dataset contains images for three separate cameras. We restrict ourselves to the Nikon D810 dataset. To this end we train a network following the aforementioned procedure by using the spectral sensitivities of the Nikon D810 camera and evaluate this network on the respective LSMI test dataset. One version of the network is used as is. Another version is fine-tuned on the corresponding LSMI training dataset, including only standard data augmentation (random rotation, zoom and crop, no pixel-level relighting [5]). The computed validation metrics are the mean and median angular errors in degrees between the predicted and actual white point maps. These are listed in Table 2.

From these results it is clear that, although our network only has about 2M parameters which is a fraction of the 56M parameters of the LSMI-U network, it performs on par if we train and validate on LSMI. When trained on simulated data, the white point angular errors increase. Looking at the failure cases for our pure simulation-trained network reveals that for some scenes in the LSMI dataset blue, green or red colored illuminants are used. Such illuminants are not present in our training database.

To further benchmark our method with respect to the LSMI-U network, we also convert our validation dataset to the Nikon D810 camera spectral sensitivities. This is possible since images from our multi-spectral camera can be transformed to mimic the spectral sensitivities of any RGB camera with adequate accuracy, see Figure 6. The resulting images are typically smoother compared to the actual images from the Nikon D810 camera. Therefore, we further fine-tuned the LSMI-U network on a smoothed version of the original Nikon LSMI training dataset. Since our dataset also includes Macbeth color charts, we can also compute color accuracy (DE2000) after white balancing.

The results in Table 3 show that our pure-simulation trained network greatly outperforms the LSMI-U network. Both in terms of white point angular error and color accuracy. It is also clear that fine-tuning our network on the LSMI training dataset is detrimental to the overall accuracy. This clearly shows that the LSMI-U network is overfit on the LSMI dataset and cannot generalize to previously unseen scenes. This is particularly clear for an exemplar outdoor scene from our validation dataset shown in Figure 7. Note that it is not directly possible to compare the results from Tables 2 and 3 as they are computed for a, respectively, dense and sparse local white point map.

As such, we can conclude that both networks perform similarly when trained and evaluated on LSMI, but both results degrade substantially when evaluated on our validation dataset (with LSMI-U degrading most). This illustrates poor generaliza-

Figure 5. *Cameras mounted on a tripod.*

The mean and median white point angular errors (degrees) on the Nikon LSMI test dataset for the LSMI-U network and our network (SPTY), both trained purely on simulations (Sim.) and fine-tuned on the Nikon LSMI training dataset.

Network	Training	Mean	Median
LSMI-U	LSMI	1.9	1.2
SPTY	LSMI	2.0	1.6
SPTY	Sim.	2.5	1.8

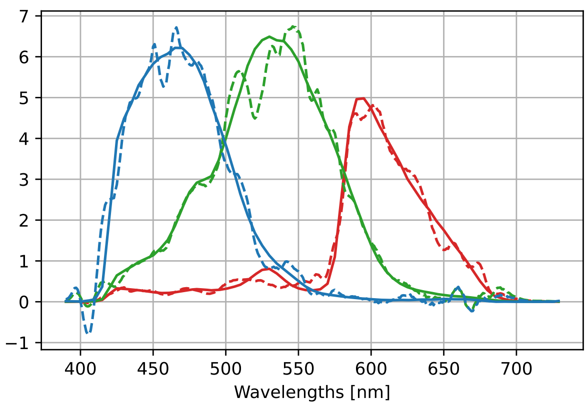

Figure 6. The true spectral sensitivities of the Nikon D810 camera (solid) and the approximation by transforming the spectral sensitivities of the S1-M VIS camera module (dashed).

The mean and median white point angular errors (degrees) and color accuracy (DE2000) on the MIMI validation dataset converted to the Nikon D810 camera for the fine-tuned LSMI-U network and our network (SPTY), both trained purely on simulations (Sim.) and fine-tuned on the Nikon LSMI training dataset.

Network	Training	Mean	Median	DE2000
LSMI-U	LSMI	4.4	3.7	6.3
SPTY	LSMI	3.8	2.7	6.0
SPTY	Sim.	3.1	2.0	5.1

Figure 7. Exemplar LSMI-U failure case of the MIMI validation dataset converted to the Nikon D810 camera. Left: the non-white balanced image. Right: the white balanced result using the LSMI-U network where nearly all color has been removed.

tion after training on a medium sized dataset. On the other hand, training on simulated data and evaluating on LSMI or on our validation dataset yields fairly consistent results. As expected, these results are worse compared to those achieved by training and validating on LSMI, but substantially better than those obtained by training on LSMI and evaluating on our validation dataset.

RGB-based vs. multi-spectral-based white point detection

To demonstrate the benefit of using multi-spectral information for local white point detection, we apply the SPTY network to our validation dataset (MIMI). Two network variants are tested, one that starts from the Samsung Galaxy Tab S9 RGB images and predicts local white points for the same camera and one that starts from the multi-spectral images and predicts local white points for the multi-spectral camera. One additional variant is tested that extends the latter by converting the multi-spectral local white points to the Samsung Galaxy Tab S9 RGB camera. Similarly to the results above, we use mean and median white point angular errors and color accuracy as metrics. To make the white point angular error metrics comparable between the multi-spectral and RGB white points, the multi-spectral white points are converted to the RGB domain prior to white point angle computation.

The mean and median white point angular errors (degrees) and color accuracy (DE2000) on the MIMI validation dataset for our networks.

Network variant	Mean	Median	DE2000
RGB → RGB	2.8	1.9	4.8
MSI → MSI	2.0	1.5	3.6
MSI → MSI → RGB	2.3	1.8	n/a

The results from Table 4 clearly illustrate that using multi-spectral image data leads to more accurate white point detection results. Converting the multi-spectral white point to RGB gives slightly worse results. This can be explained from small cross-camera differences and an imperfect conversion, especially for fluorescent illuminants. Also the high color accuracy of the fully multi-spectral method becomes apparent. This is due to the fact that removing the illuminant from the multi-spectral measurements using the white point is much more accurate compared to RGB.

Conclusion

In this paper we present a local white point detection method, leveraging compact multi-spectral camera technology. We show that capturing additional, narrow-band spectral information beyond the RGB channels yields more accurate white point estimates and improved color accuracy. Specifically, we built a novel neural network-based algorithm that performs on par with state-of-the-art methods when trained on RGB data, but has significantly less parameters and is shown to generalize better. Based on results drawn from a new dataset, recorded with our S1-M VIS snapshot camera module we observe improved white point angles and color accuracy.

Whereas we report the final algorithm performance based on measured data, we trained our neural networks on simulated measurements to mitigate the required training set size. The new dataset is broadly applicable for the comparison of cameras, as it can be converted accurately to any RGB camera. The multi-illuminant multi-spectral image (MIMI) dataset will be made publicly available.

We plan to extend the MIMI validation dataset and devise a measurement protocol to obtain dense local white point information applicable in-the-wild. Another possible direction for future work is to explicitly model the local white point ambiguity using the multi-spectral information and to rely on complementary information (e.g. scene content) to select the actual local white points.

References

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

[2] Dongliang Cheng, Dilip K. Prasad, and Michael 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, April 2014

[3] Jonathan T. Barron. Convolutional Color Constancy, ICCV 2015

[4] Yuanming Hu, Baoyuan Wang, and Stephen Lin. FC 4: Fully Convolutional Color Constancy with Confidence-weighted Pooling, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4085–4094, 2017

[5] Dongyoung Kim, Jinwoo Kim, Seonghyeon Nam, Dongwoo Lee, Yeonkyung Lee, Nahyup Kang, Hyong-Euk Lee, ByungIn Yoo, Jae-Joon Han, and Seon Joo Kim. Large Scale Multi-Illuminant (LSMI) Dataset for Developing White Balance Algorithm Under Mixed Illumination Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 2410–2419, 2021.

[6] Dongyoung Kim, Jinwoo Kim, Junsang Yu, and Seon Joo Kim. Attentive Illumination Decomposition Model for Multi-Illuminant White Balancing, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 25512–25521, 2024

[7] Wenjun Wei, Yanlin Qian, Huaian Chen, Junkang Dai, and Yi Jin. Integral Fast Fourier Color Constancy, IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2025

[8] Zhuo Hui, Aswin C. Sankaranarayanan, Kalyan Sunkavalli, amd Sunil Hadap. White balance under mixed illumination using flash photography, IEEE International Conference on Computational Photography (ICCP), 2016

[9] Yinqiang Zheng, Imari Sato, and Yoichi Sato. Illumination and reflectance spectra separation of a hyperspectral image meets low-rank matrix factorization, Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2015), pp. 1779–1787, 2015

[10] Samu Koskinen, Erman Acar, and Joni-Kristian Kämäräinen. Single Pixel Spectral Color Constancy. Int J Comput Vis 132, pp. 287–299 (2024).

[11] Ortal Glatt, Yotam Ater, Woo-Shik Kim, Shira Werman, Oded Berby, Yael Zini, Shay Zelinger, Sangyoon Lee, Heejin Choi, and Evgeny Soloveichik. Beyond RGB: A Real World Dataset for multispectral Imaging in Mobile Devices, Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), pp. 4344-4354, 2024

[12] Juho Lee, Yoonho Lee, Jungtaek Kim, Adam Kosiorek, Seungjin Choi, and Yee Whye Teh. Set Transformer: A Framework for Attention-based Permutation-Invariant Neural Networks, Proceedings of the 36th International Conference on Machine Learning, pp. 3744–3753, 2019