Beyond RGB: A Real World Dataset for Multispectral Imaging in Mobile Devices

Ortal Glatt[1]   Yotam Ater[1]   Woo-Shik Kim[2]   Shira Werman[1]   Oded Berby[1]

Yael Zini[1]   Shay Zelinger[1]   Sangyoon Lee[2]   Heejin Choi[2]   Evgeny Soloveichik[1]

Samsung Israel R&D Center[1], Israel {ortal.glatt,yotam.ater}@samsung.com
Samsung Advanced Institute of Technology[2], Korea wooshik.kim@samsung.com

Abstract

Multispectral (MS) imaging systems have a wide range of applications for computer vision and computational photography tasks, but do not yet enjoy widespread adoption due to their prohibitive costs. Recently, advances in the design and fabrication of photonic metamaterials have enabled the development of MS sensors suitable for integration into consumer grade mobile devices. Augmenting existing RGB cameras and their processing algorithms with richer spectral information has the potential to be utilized in many steps of the image processing pipeline, but diverse real world datasets suitable for conducting such research are not freely available. We introduce Beyond RGB[1], a real-world dataset comprising thousands of multispectral and RGB images in diverse real world and lab conditions that is suitable for the development and evaluation of algorithms utilizing multispectral and RGB data. All the scenes in our dataset include a colorimetric reference and a measurement of the spectrum of the scene illuminant. Additionally, we demonstrate the practical use of our dataset through the introduction of a novel illuminant spectral estimation (ISE) algorithm. Our algorithm surpasses the current state-of-the-art (SoTA) by up to 58.6% on established benchmarks and sets a baseline performance on our own dataset.

1. Introduction

Cameras require color filter arrays (CFAs) [43] to differentiate between different wavelengths of light and allow rendering of colorful images. The ubiquitous Bayer array [8] introduced in 1976 uses three broad band filters (red, green and blue) in a manner that emulates the color perception of the human eye [28]. Despite their widespread application, an inherent drawback to the employment of wide band filters is the loss of spectral information from the original optical signal which carries significant poten-

tial for a wide range of computer vision applications. These include, but are not limited to, color and material analysis, agricultural inspection, food quality control and healthcare [1, 33, 42, 54].

Traditional hyperspectral (HS) and multispectral (MS) imaging systems, such as push broom, line scan or tunable filter cameras, have the ability to capture images with high spectral resolution [3, 11, 45]. However, widespread adoption of traditional HS and MS systems is unfeasible due to their prohibitive cost, large size, long image acquisition times and the need for substantial operator expertise. Recent advances in the design of nanophotonic elements have facilitated the creation of an MS imaging sensor in a compact and low cost form factor suitable for mobile devices by fabricating a CFA with multiple spectral channels onto a consumer grade CMOS image sensor [38, 59].

Recreational photography is by far the most common use case of imaging systems today, so it is natural to explore how a mobile device augmented with an MS sensor could benefit from additional spectral information. Previous work has shown that MS data can be utilized to improve illuminant estimation [64], to extract an estimate of the spectrum of a global scene illuminant [62, 71] or an estimate of the spectrum of multiple local illuminants [40, 57]. Additionally, in [17] it was shown that RGB cameras can be made more colorimetric by utilizing carefully chosen external filters so that the overall system response more closely obeys the Luther condition [25, 30, 49]. These results indicate that there is a potential for harnessing added spectral information to improve both the auto white balancing (AWB) and color space conversion processing steps which are crucial to rendering colorimetrically accurate and visually pleasing images [15, 31].

Developing algorithms which utilize both MS and RGB images requires a diverse and large-scale dataset collected in the real world with both MS and RGB cameras, accompanied by suitable ground truth information. Despite the availability of numerous HS and spectral reflectance (SR) datasets [4, 10, 18, 20, 40, 50–52, 60, 65, 70], they are often

---

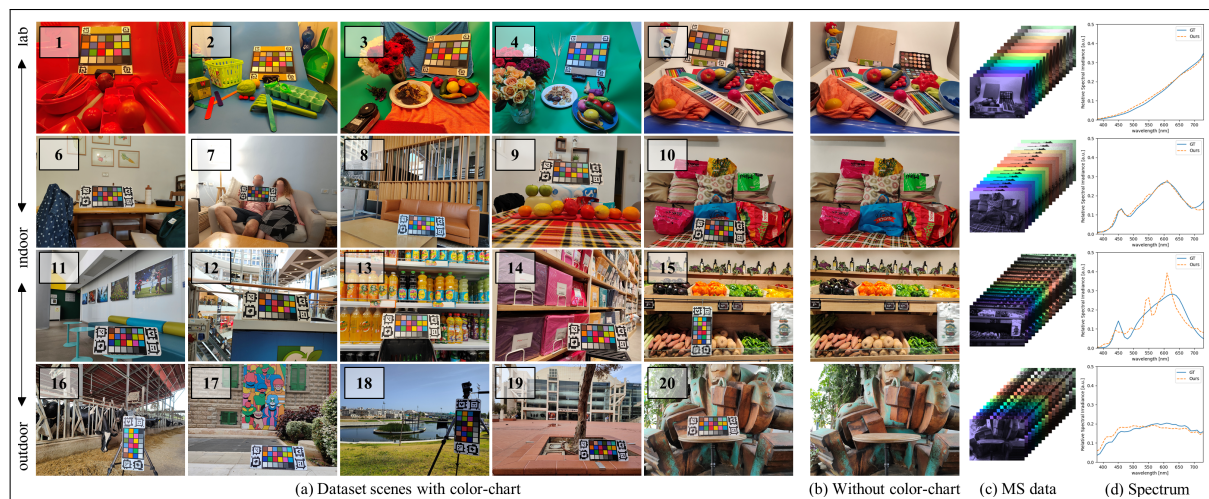[1]Dataset available at: https://github.com/shirawerman/Beyond-RGB

Figure 1. Illustrative examples from our dataset: (a) 1-5 sample scenes captured in a viewing booth, 6-15 sample scenes captured in the field in indoor locations, 16-20 sample scenes captured in the field in outdoor locations; (b) the counterpart samples of 5, 10, 15 and 20 captured without the color chart; (c) the multispectral image cube of scenes 5, 10, 15 and 20; (d) the groundtruth illuminant spectrum measurement of scenes 5, 10, 15 and 20 together with the estimation performed by our algorithm.

limited in size and content diversity which limits their utility for algorithm development. To bridge the data gap, we introduce the Beyond RGB dataset which includes 1680 diverse scenes collected in the lab and in the field, samples of which are shown in Figure 1. We discuss the comparative differences between datasets in Section 2.

To demonstrate the utility of Beyond RGB, we introduce an adaptation of the well known Convolutional Color Constancy (CCC) [5] algorithm, which we modify to perform the illuminant spectral estimation (ISE) task. Our algorithm achieves SoTA performance on previous ISE benchmarks and provides a baseline performance level on our own dataset.

Our main contributions can be summarized as follows: (1) A diverse and large scale dataset of RGB and MS images from mobile devices with both colorimetric and illuminant spectrum ground truth, suitable for ISE, color constancy and color accuracy tasks. (2) An adaptation of CCC to the ISE task that improves on existing SoTA by up to 58.6%. (3) We show that our dataset and ISE algorithm are easily adapted for fusion between MS and RGB data from different devices.

2. Related Work

2.1. Related Datasets

Numerous HS and SR datasets have been developed over the years [4, 10, 18, 20, 40, 50–52, 60, 70] (Table 1). The distinction between the MS and HS domains is not clear cut and usage of the terms varies across the literature. However, we observe that generally when the spectral bandwidth
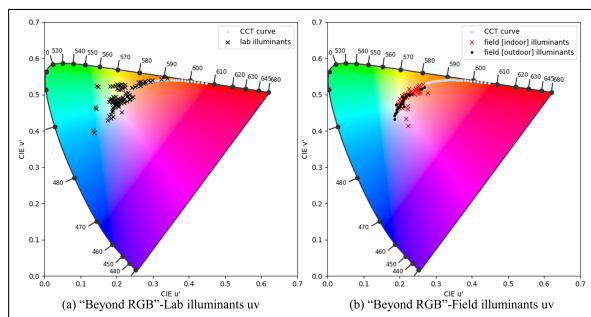

Figure 2. CIE-1976-Luv uv chromaticity coordinates of illuminants in the Beyond RGB dataset: (a) illuminants captured in lab scenes (b) illuminants captured in the field. The dataset contains 93, 43 and 29 different illuminants in lab, indoor and outdoor scenes respectively.

of each channel is less than 10nm or, equivalently, the instrument has more than 30 channels in the visible region, it is considered to be HS. HS image datasets are convenient for research, because high resolution spectral cubes can be post-processed to include a subset of the original channels as required for the application of interest. It is for this reason that typically HS datasets are collected, as can be seen in Table 1, as they allow for more versatile post-processing than MS datasets which contain less spectral information.

If the scene illuminant is known, usually through inclusion of an achromatic reference target in the scene, HS datasets may be post processed so that the scene illuminant is removed from the HS image leaving an SR image. The

Name	Data	Scenes	Indoor	Outdoor	Staged	Spectral Range (Channels)	Colorimetric Reference?	Illuminant Spectra?
NS-FOS1 [51]	HS	30	0	30	0	400-720nm (32)	No	Yes
NS-FOS2 [20]	SR	8	1	7	0	400-720nm (32)	No	Yes
NS-FOS3 [19]	HS	36	0	36	0	400-720nm (32)	No	Yes
NS-FOS4 [50]	HS	30	0	30	0	400-720nm (32)	No	Yes
NS-FOS5 [18, 21]	SR	50	0	50	0	400-720nm (32)	No	Yes
CAVE [70]	SR	32	0	0	32	400-700nm (31)	Yes	Yes
Scyllarus [60]	HS	73	13	45	15	400-720nm (32)	No	No
NUS [52]	HS	66	8	27	31	400-700nm (30)	No	Yes
Harvard [10]	HS	75	36	39	0	420-720nm (30)	No	No
ISNL [62]	HS	17	11	6	0	405-700nm (30)	Yes	Yes
KAUST [40]	SR	409	~280	~80	~49	400-730nm (34)	No	Yes
Arad 1K [4]	HS	1000	~10	~990	0	400-1000 (201)	No	No
Color-Checker [22]	RGB	568	~232	~336	0	-	Yes	No
NUS-8 [13]	RGB	1736	~543	~1193	0	-	Yes	No
Proposed	**MS+RGB**	**1680**	**235**	**237**	**1208**	**380-730nm (16)**	**Yes**	**Yes**

Table 1. A comparison of Beyond RGB and existing datasets. MS, HS, SR and RGB respectively indicate multispectral, hyperspectral, reflectance and red-green-blue type data. The colorimetric reference and illuminant spectra columns indicate whether the dataset provides the relevant ground truth source. We indicate the split between indoor, outdoor and staged scenes. We consider a scene to be "staged" if it is repeated multiple times under different illuminants, or if it is a lab setup. $\sim$ indicates an estimated number as the scene type is not always clear.

utility of SR datasets lies in the ability to relight scenes in post-processing with any desired illuminant. The nominal domain on which an ISE algorithm is required to operate is defined by the space of all possible scenes multiplied by the space of all possible illuminants. Therefore, relighting SR data reduces the requirement to capture many scenes under many different lighting conditions. However, as observed in [23], the naive simulation of illuminants by multiplying an SR image with an illuminant spectra does not generate real-world effects such as inter-reflections, specularity, shading or variation in intensity across the scene. Additionally, there exists a large class of image regions which are not well modelled by SR maps. Important examples of these cases are specular highlights and blue sky, which also serves as a memory color [56] and is thus particularly important to be rendered correctly.

Our dataset contains real world images, including regions which are often avoided in SR datasets, and since the distribution of lighting in Beyond RGB is not post-processed synthetically, it is guaranteed to be physically accurate. We empirically demonstrate the existence of a domain gap between a relit SR dataset and our own dataset in Section S2.2 of the supplementary material.

A second limitation of HS and SR datasets is the long exposure time required for image acquisition, due to the mechanical scanning mechanisms and narrow bandwidth filters often used by HS imagers. This complication makes it difficult to photograph moving subjects (people, animals, etc.) and is time consuming which leads to smaller datasets

as can be seen in Table 1, where the majority of datasets have a bias to outdoors images and only [4, 40] contain more than a few tens of images. To further emphasize, we note that with the exception of a handful of images in [60, 70] none of the HS and SR datasets surveyed in Table 1 contain any images of people; 11% of the field scenes in our own dataset contain people or animals. It is due to these limitations that Gijseni et al. recommend utilizing both SR and real world datasets for the evaluation of color constancy algorithms [23]. Indeed, the two datasets which have become most commonly used in color constancy research are the real world Color-Checker [22] and NUS-8 [13] datasets and we structure our own dataset after these.

Of the available HS and SR datasets the two most similar to our own are Arad 1K [4] and KAUST [40] datasets which have a similar amount of scenes. While Arad 1K has a large amount of diverse field scenes, it is captured almost exclusively outdoors and does not contain a colorimetric reference nor is the illuminant spectra measured making it unsuitable for the tasks we are interested in. The KAUST dataset is similar in magnitude to our own, however due to the limitations mentioned above it is limited in the diversity of subjects it captures.

2.2. Illuminant Spectral Estimation

The image formation model giving rise to the ISE problem for a camera with N channels is,

$$\mathbf{I}_k(x, y) = \int_\Lambda R(x, y; \lambda) L(\lambda) f_k(\lambda) d\lambda \quad k \in [0..N] \quad (1)$$

Where (x, y) are the spatial coordinates of a given pixel, k is the channel index, $f_k(\lambda)$ is the system response function for the k-th channel, $\mathbf{I}_k$ is the k-th channel of the observed image, and λ is the parameter indicating the observed wavelength of light. R and L are respectively the unknown scene reflectance and the unknown scene illuminant. Λ is the domain of integration over wavelengths where the camera has non-negligble response. The model of Eq. 1 is a common simplification that assumes a spatially uniform illuminant and negligble specular highlights.

Our goal is to estimate $L(\lambda)$ given only the observed image $\mathbf{I}(x, y)$, a task that is well known to be fundamentally under-determined. Overcoming the under-determined nature of the problem is the central challenge in ISE and previous works have addressed it with subspace modelling of both illuminance and reflectance [12, 24], and by careful formulation of the optimization problem and it's heuristics [16, 29]. Another common approach is to exploit the low dimensionality of natural reflectance spectra [44, 55] and to pose the problem as a low rank matrix factorization problem [71]. Other strategies involve selection of pixels which are likely to be achromatic [9, 37, 62, 66] and thus reflect the required spectra without modification. Finally, recent approaches have utilized deep learning techniques to perform pixelwise ISE utilizing a CNN [58] and by explicitly modeling matrix factorization into an unrolling network [40]. However, the number of learning based approaches have so far been limited by the lack of suitable data, which was unavailable prior to the release of the KAUST dataset by Li et. al. [40].

Color constancy, the task of estimating the color of the scene illuminant, is a widely studied topic [2, 5, 7, 9, 26, 27, 37, 39, 41, 61, 66, 68, 69] due to it's central role in rendering correct color in images. It is closely related to the ISE task, solving the same fundamental Eq. 1 but without seeking to explicitly recover the full spectra. In CCC [5, 7] Barron et. al. formulated the log-chroma histogram feature, which has been widely influential in further work [35, 69], and posed the color-constancy problem as a discriminative learning task. More recently, deep learning based methods have made large performance improvements on the color constancy task by applying contrastive learning, deep metric learning and the transformer architecture [39, 41, 69].

Taking our cue from the progression of techniques demonstrated to be effective on the color constancy problem, and with the availability of suitable datasets such as our own, we show that the core ideas of [5] can be transferred to the ISE task with significant performance gains compared to previous methods. We hope the community will utilize Beyond RGB to further this avenue of research.

3. Beyond RGB Dataset

Scenes. Beyond RGB contains 1680 scenes, of which 472 are collected in the field and 1208 are collected with an automated viewing booth. The scenes collected in the lab, are comprised of 13 distinct scenarios a sample of which can be seen in the top row of Figure 1. We designed these scenarios, so that some include a wide gamut of colors (e.g. Figure 1.a.5), while others have a few primary dominant colors (e.g. Figure 1.a.1). The automated lab setup is a custom built viewing booth, with 12 different light sources which can be turned on and off using a controller. The illumination in the viewing booth is spatially uniform, and the ability to control light sources allows us to generate many distinct mixtures. In practice we generate 93 different illuminations, whose location on the CIE1976 chromaticity diagram is shown in Figure 2.a. The lab illuminants provide good coverage of the Planckian locus from a CCT [47] of 1801K to a CCT of 14132K, and also include some heavily tinted illuminants off of the locus.

The field scenes allow us to capture real world scenarios and are comprised of 235 scenarios in indoors settings and 237 scenarios in outdoors settings. They are collected at many different locations, hours of the day and contain diverse subject matter such as people, animals, buildings and vegetation. The illuminants present in the field scene are primarily located on the Planckian locus (Figure 2.b) as can be expected for natural light. The field illuminants also have good coverage of the Planckian locus spanning a CCT range of 2642K to 7278K in indoor scenes, and 4021K to 11753K in outdoor scenes.

MS Sensor. For MS images, we use a prototype camera with 16 distinct filters arranged in a 4x4 CFA shown in Figure S1, and a native resolution of 2584x1936 pixels which corresponds to a resolution of 646x484 pixels in each spectral channel. The Fabry-Perot type bandpass filters are fabricated and placed on top of pixels of a CMOS image sensor (CIS). We provide the MS data as an HDF5 [34] file in the native resolution with minimal post-processing that consists of removing the black level, normalizing by the saturation level, correcting bad pixels by linear interpolation with nearest neighbors of the same channel, correcting lens shading and noise reduction with an edge preserving bi-lateral filter [48].

RGB Cameras. For reference RGB cameras we utilize two widely available flagship Android phones. A Samsung Galaxy S21 Plus SM-G996B with a native resolution of 4032x3024 pixels and an Oppo Find X5 Pro CPH2305 with a native resolution of 4096x3072 pixels. We furnish the original RAW image data as an HDF5 file along with the corresponding metadata, only modified as necessary to ensure privacy. The RGB images are captured simultaneously with the MS image to ensure that the same scene is viewed in both images.

Camera Intrinsics and Extrinsics. For both the RGB and MS cameras we provide camera intrinsics in the form of a camera projection matrix and distortion parameters. To allow for rectification of scenes we also provide the relative pose of the RGB and MS cameras in scenes where the color-chart is presented. Further information is provided in section S3.2 of the supplementary material.

Spectrophotometer. We acquire spectral measurements using an X-Rite i1Pro 3 Plus in ambient mode, which utilizes a calibrated cosine corrected diffusive element to collect half-hemisphere incident illumination. The spectrophotometer collects spectra in the range of 380-730nm at a resolution of 10nm. Each spectra is measured near the color calibration target with the sensor head facing away from the target.

Color Calibration. To characterize color accuracy we include an X-Rite ColorChecker Classic calibration board in every scene. We frame the ColorChecker with April-Tag fiducial markers [53, 67] for accurate detection in all scenes, and provide bounding contours for the patches of the color chart in all images. To avoid the need for masking or cropping off the color chart when the dataset is used for algorithm training, we photograph each scene in the dataset twice: once with the color chart presented, and then once again with the color chart removed.

Privacy. In accordance with the privacy protocol established in [36], we blur faces and license plates present in the dataset. An H/3 x H/3 average blurring window was applied separately to each channel of RGB or MS image, where H is the height of the blurred area, thus maintaining the average signal level originally present in the region. The blurred area edges includes outer border pixels for the average calculation. We provide coordinates of regions that have been blurred.

4. Illuminant Spectral Estimation Method

Our algorithm is comprised of three main stages: (1) Division of an input MS image into triplets, (2) processing of each triplet by a CCC block and (3) converting the CCC block outputs to a spectrum estimate using a CNN. An overview of our method is shown in Figure 3. Implementation details can be found in Section S1 of the supplementary material.

4.1. Multispectral Triplets

Given a raw MS image $\mathbf{I}_{\text{raw}}$ of shape $H \times W$ patterned according to the CFA in Figure S1, we rearrange it into an MS cube of shape $16 \times \frac{H}{4} \times \frac{W}{4}$ which we denote as $\mathbf{I}$. From the 16 channels, we choose M different triplets T_i:

$$T_i = [c_i^1, c_i^2, c_i^3] \qquad (2)$$
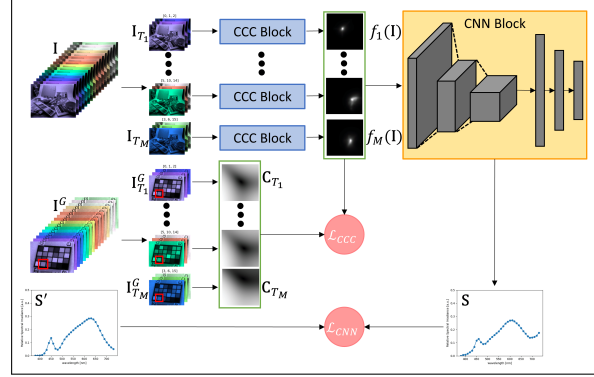
where the c_i^k are channel indices in $\{0, ..., 15\}$.

Figure 3. Method description: **Top** The channels of an input MS cube **I** are divided into M triplets, each of which is converted to a log-chroma histogram and processed by a CCC block. The CCC blocks output score maps $f_i(\mathbf{I})$, which are then concatenated and processed by a CNN which outputs the estimated spectrum **S**. **Bottom** During training we use the colorimetric reference in $\mathbf{I}^G$ to extract GT triplet ratios and generate M cost maps $\mathbf{C}_{T_i}$ which are used for calculating $\mathcal{L}_{\text{CCC}}$. $\mathcal{L}_{\text{CNN}}$ is calculated using the measured GT spectrum $\mathbf{S}'$.

For each triplet T_i we create a stack of the relevant channels from the MS cube **I**:

$$\mathbf{I}_{T_i} = [\mathbf{I}_{c_i^1}, \mathbf{I}_{c_i^2}, \mathbf{I}_{c_i^3}] \qquad (3)$$

Concretely, we use $M = 21$ and choose 14 triplets of consecutive MS channels and 7 additional triplets from channels which have little spectral overlap.

4.2. Single CCC Block

The CCC block and loss used in our method are implemented exactly as defined in [5]. For clarity in explaining our method we repeat some of the definitions, but we refer the reader to the original paper for a full discussion.

For each channel triplet T_i we construct the corresponding normalized 2D log-chroma histogram $\mathbf{N}_i(u, v)$. The log-chroma histogram is convolved with a learned filter using the "pyramid context" technique [6], by constructing a pyramid from the histogram, convolving each scale of the pyramid with a small learned filter, and then collapsing the filtered histogram. For notational simplicity, let us denote the convolutions pyramid as a single convolutional filter $\mathbf{F}_i$ for triplet T_i as:

$$f_i(\mathbf{I}) = \text{softmax}(\mathbf{N}_i * \mathbf{F}_i) \qquad (4)$$

The filters $\mathbf{F}_i$ are learned using the loss function defined in CCC. First, we define a cost map $\mathbf{C}$ for triplet T_i as:

$$\mathbf{C}_{T_i}(u, v) = \text{AE}\left(\begin{bmatrix} e^{-u_i'} \\ 1 \\ e^{-v_i'} \end{bmatrix}, \begin{bmatrix} e^{-u} \\ 1 \\ e^{-v} \end{bmatrix} \right) \qquad (5)$$

Where u, v and u'_i, v'_i are respectively the observed and ground truth chrominance values for triplet T_i, the exponent is used in order to cancel the log from the definition of uv, and the angular error between two vectors is defined as:

$$\text{AE}(\mathbf{v}_1, \mathbf{v}_2) = \arccos\left(\frac{\langle \mathbf{v}_1, \mathbf{v}_2 \rangle}{\|\mathbf{v}_1\| \|\mathbf{v}_2\|}\right) \qquad (6)$$

The final CCC loss function is the multiplication between the uv score maps $f_i(\mathbf{I})[(u, v)]$, and the uv cost map calculated using the log-chrominance GT:

$$\mathcal{L}^i_{\text{CCC}} = \sum_{u,v} f_i(\mathbf{I})(u, v) \cdot \mathbf{C}_{T_i}(u, v) \qquad (7)$$

4.3. Multiple CCC Blocks

We learn a separate CCC block for each triplet T_i. In each block, we use a 7-level pyramid comprised of 5x5 convolution kernels and bilinear interpolation for the up-sampling and down-sampling processes.

The spectral information of neutral areas in a given scene, is correlated with the scene illuminant spectra. To calculate GT chrominance values for each triplet, the illuminant value of each MS channel is needed. In order to get the most precise value of the illuminant for each channel, we use the colorimetric reference available in our dataset. We use the median of a gray patch in $\mathbf{I}^G$ as shown in Figure 3, and calculate:

$$\begin{aligned} u'_i &= \log(\text{med}(\mathbf{I}^G_{c^2_i})/\text{med}(\mathbf{I}^G_{c^1_i})) \\ v'_i &= \log(\text{med}(\mathbf{I}^G_{c^2_i})/\text{med}(\mathbf{I}^G_{c^3_i})) \end{aligned} \qquad (8)$$

Where $\mathbf{I}^G$ is the gray patch MS data in the MS image with the color-chart presented. The cost map $\mathbf{C}_{T_i}$ defined in Eq. 5 is calculated for each triplet using u'_i, v'_i, and then used to calculate the loss component defined in Eq. 7. Each CCC block output is effectively a score map for the log-chroma histogram bins, so we expect to get a higher score for uv bins which are close to the true uv of each relevant triplet.

4.4. CNN Block

The (u, v) spatial distribution of the score maps contains information which we want to utilize to estimate the scene illuminant. We concatenate the scores maps, and input to a CNN block that is trained to predict the relative power spectral density $\mathbf{S}$.

The loss for the CNN is taken between the predicted spectrum S and measured GT spectrum S',

$$\mathcal{L}_{\text{CNN}} = \text{AE}(\mathbf{S}', \mathbf{S}) \qquad (9)$$

In the case of the Beyond RGB dataset, $\mathbf{S}'$ is a 36 channel vector representing the relative power spectral density

between 380nm to 730nm at 10nm intervals, measured using the spectrophotometer as described in Section 3.

The full training loss used is,

$$\mathcal{L} = w_1 \cdot \sum_{i \in [0,M]} \mathcal{L}^i_{\text{CCC}} + w_2 \cdot \mathcal{L}_{\text{CNN}} + w_3 \cdot \mathcal{L}_{\text{wd}} \qquad (10)$$

Where $\mathcal{L}_{\text{wd}}$ is a weight-decay regularization component and $w_{1,2,3}$ are scalar weights to balance the loss contributions. The gradients from $\mathcal{L}_{\text{CNN}}$ propagate only to the CNN block, and are not propagated back to the CCC blocks.

5. Experiments & Results

We evaluate the proposed method on two established reflectance datasets KAUST and CAVE [40, 70] which we relight using 59 standard illuminants from [46], and on our own Beyond RGB dataset.

We report the commonly used angular error metric of Eq. 6 in degrees. However, since some of the compared methods output results which are not equivalent to a full spectrum estimation, we report three flavors of angular error:

$$\begin{aligned} \Delta\text{A}_{\text{HS}} &= \text{AE}(\mathbf{S}', \mathbf{S}) \\ \Delta\text{A}_{\text{MS}} &= \text{AE}(\mathbf{M}_{\text{CAL}}\mathbf{S}', \mathbf{S}_{\text{MS}}) \\ \Delta\text{A}_{\text{XYZ}} &= \text{AE}(\mathbf{M}_{\text{CMF}}\mathbf{S}', \mathbf{S}_{\text{XYZ}}) \end{aligned} \qquad (11)$$

Where $\mathbf{S}'$ is the GT spectrum measurement and $\mathbf{S}$ is the algorithms' full spectrum estimate. $\mathbf{S}_{\text{MS}}$ is the output estimate for algorithms which output their illuminant estimate in the MS channel space and similarly $\mathbf{S}_{\text{XYZ}}$ denotes output in the CIE-1931 XYZ space. $\mathbf{M}_{\text{CAL}}$ is the calibration matrix converting $\mathbf{S}'$ to an expected 16 channel MS signal derived from the MS sensor filter responses (Fig. S1), and $\mathbf{M}_{\text{CMF}}$ is the standard CIE-1931 CMF [14] which converts $\mathbf{S}'$ to the XYZ space.

5.1. Spectral Reflectance Datasets

We work with KAUST and CAVE SR images of shape 31x512x512, that represent the SR of each pixel at wavelengths between 400nm and 700nm in 10nm intervals. The KAUST dataset has SR information up to 730nm, but we truncate the last few channels so that the data is equivalent to CAVE. We relight the scene by multiplying each pixels' reflectance vector by a diagonal matrix with the spectrum of the desired illuminant to get a 31 channel HS image $\mathbf{I}_{31}$. To simulate our own 16 channel MS sensor, we also convert $\mathbf{I}_{31}$ into a 16 channel MS measurement using the calibration matrix $\mathbf{M}_{\text{CAL}}$ derived from the MS channel responses of Fig S1 for a final output $\mathbf{I}_{16}$ of size 16x512x512. For both datasets, we mask the white reference while training and testing.

We train our proposed method using both the full HS $\mathbf{I}_{31}$ input and the lower dimension MS $\mathbf{I}_{16}$ input to perform a

Dataset	Method	mean-ΔA_{HS} ↓	std-ΔA_{HS} ↓
KAUST	GrayEdge	11.46	8.02
	LRMF	24.06	23.49
	ISNL	20.05	11.46
	PWIR	28.07	21.20
	DUN	9.17	6.30
	Ours-16ch	8.98	4.97
	Ours-31ch	7.30	4.45
CAVE	GrayEdge	18.33	14.89
	LRMF	20.05	15.47
	ISNL	22.92	16.62
	PWIR	34.95	21.20
	DUN	16.04	13.75
	Ours-16ch	9.46	5.05
	Ours-31ch	6.64	3.62

Table 2. ΔA_{HS} comparison between the proposed method using HS data (Ours-31ch), the proposed method using MS sensor raw space data (Ours-16ch), and other ISE methods, on KAUST & CAVE datasets. Results of other methods on KAUST and CAVE are as reported in [40]. Green and yellow highlights respectively indicate best and second best results.

Dataset	Method	ΔA_{HS} ↓		ΔA_{MS} ↓	
		mean	std	mean	std
Lab	GrayWorld			6.17	4.09
	GrayEdge			8.97	5.02
	LRMF			21.59	2.28
	PWIR	27.10	10.86	11.03	6.81
	Ours	5.92	2.92	2.05	1.39
Field	GrayWorld			6.85	2.40
	GrayEdge			11.39	3.93
	LRMF			21.90	1.20
	PWIR	16.31	6.07	6.07	3.35
	Ours	7.22	5.54	2.73	2.46

Table 3. ΔA_{HS} and ΔA_{MS} comparison between the proposed method and other ISE methods on the Beyond RGB dataset. Green and yellow highlights respectively indicate best and second best results.

fair comparison to other methods which utilized the full HS data. When training on full HS input we select 29 consecutive triplets as inputs. In both cases training is done on random 256x256 crops of the images. We follow the protocol of [40] and train on a random 80% split of HS images from KAUST with 80% of the illuminants, and report the evaluation results on the remaining 20% of images and illuminants.

We compared our proposed method with six existing methods for illuminant spectral estimation, including GrayEdge [32, 66], LRMF [71], ISNL [63], PWIR [57] and DUN [40], using the results reported by the last. As shown in Table 2, our proposed method using the full 31 HS channel input achieves the best results on both datasets outperforming DUN by 58.6% and 20.3% on CAVE and KAUST respectively. Utilizing only the 16 channel MS input degrades the results but still outperforms previous methods which utilized the full HS input by 41% and 2%. On the CAVE dataset, which is not included in training, the proposed method generalizes significantly better than the compared methods.

5.2. Beyond RGB Dataset

Beyond RGB dataset evaluation is performed on a randomly selected 80%/10%/10% train/val/test split, with the split being taken separately between lab and field scenes to avoid over representation of lab scenes in the 10% splits. The model is trained on random 256x256 crops of the scenes without the colorimetric reference. We evaluate the performance of the proposed method on Beyond RGB

data compared to GrayWorld (GW) [9, 32], GrayEdge (GE), LRMF and PWIR. We could not evaluate DUN on the Beyond RGB dataset, since it requires reflectance data for training.

Table 3 shows statistics of ΔA_{HS} and ΔA_{MS} on the Beyond RGB lab and field scenes, and sample results are shown in Figure 4. GrayWorld, GrayEdge and LRMF are evaluated using only ΔA_{MS}, because their natural output on Beyond RGB is a 16 channel vector in the MS sensor raw space and direct conversion from $\mathbf{S}_{MS}$ to $\mathbf{S}$ leads to large errors due to the spectral response of our real-world filters (Fig. S1) which are not as clean and narrow as the responses typically found in HS data.

We observe that the magnitude of ΔA_{HS} of the proposed method on Beyond RGB is similar to that achieved when evaluating on KAUST and CAVE datasets in Table 2, even though it utilizes a real world MS sensor suitable for mobile devices. This result, together with the SoTA performance achieved on SR datasets even with reduced channels, emphasizes the potential of solving the ISE problem using MS sensors on mobile devices. We note that PWIR performs worse on the lab images than on the field images, which may indicate that it has difficulty generalizing to illuminants that are far from the Planckian locus.

5.3. MS & RGB Fusion

Utilizing the RGB data in Beyond RGB, we demonstrate a simple extension of our proposed method to include RGB and MS data from different devices. We train our method with three different input modalities: (1) Only MS data as explained in Section 4 and Section 5.2. (2) Only RGB data by utilizing RGB images as a single triplet input to our method. And, (3) fusion of MS and RGB data by expanding our standard MS algorithm with an additional triplet from
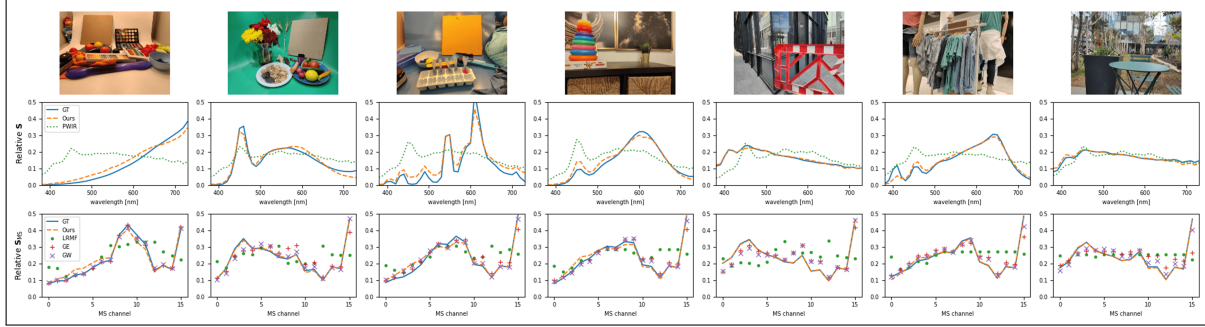
Figure 4. Sample results from the Beyond RGB test set: **Top** RGB image of the scene. **Middle** Spectrum estimate of our method and PWIR compared to the GT. **Bottom** MS channel estimates of our, LRMF, GE and GW methods compared to the GT.

Dataset	Version	$\Delta A_{HS} \downarrow$		$\Delta A_{XYZ} \downarrow$	
		mean	std	mean	std
Lab	RGB-O			7.81	7.94
	RGB-S			6.39	7.29
	MS	5.84	3.46	2.30	2.08
	Fusion-O	5.54	3.23	2.26	2.15
	Fusion-S	5.72	3.38	2.26	2.05
Field	RGB-O			6.08	4.66
	RGB-S			6.76	6.87
	MS	7.44	4.51	3.44	2.75
	Fusion-O	7.12	4.73	3.18	2.89
	Fusion-S	7.26	5.01	3.28	3.09

Table 4. ΔA_{HS} and ΔA_{XYZ} comparison of different versions of the proposed method on the Beyond RGB dataset. RGB-O and RGB-S respectively indicate RGB version using Oppo and Samsung devices data. Fusion-O and Fusion-S respectively indicate fusion version of MS data together with Oppo and Samsung devices RGB data. Green and yellow highlights respectively indicate best and second best results.

an RGB image. For RGB data, we use the raw RGB image captured by the Oppo Find X5 Pro CPH2305 or Samsung Galaxy S21 Plus SM-G996B devices as an input. All the evaluations are performed with the same 80%/10%/10% train/val/test split. The output of the MS and fusion versions is our standard **S**, while that of the RGB version is trained to output $\mathbf{S}_{XYZ}$.

Table 4 shows the ΔA_{HS} and ΔA_{XYZ} statistics on the Beyond RGB dataset, using the different versions of the proposed method. The ΔA_{XYZ} results clearly show the benefit of utilizing MS data compared to RGB. In addition, the results of the MS and RGB fusion, slightly but consistently outperforms that of the MS version, which shows the potential of MS and RGB fusion based ISE solution.

6. Summary

Limitations: In contrast to HS and SR datasets, Beyond RGB is constrained to a specific set of spectral response filters. It is therefore impractical to use Beyond RGB for testing optimal filters or calibrating algorithms for spectral responses other than those used when collecting the data. Additionally, the cameras used have not been photometrically calibrated and thus Beyond RGB cannot be used for extracting absolute spectra.

Future Work: We expect future work can explore the utilization of MS data in improvement of color accuracy of RGB cameras and improvement of white-balance algorithms whose global illumination estimation step is closely related to ISE. Further work can also explore the application of techniques and computational frameworks which have been shown to be effective in color constancy algorithms, and utilize the presence of multiple cameras in the dataset to assess cross-camera MS/RGB algorithms. It would also be advantageous to understand the source of the domain gap between re-lit SR data and directly captured MS data as in Beyond RGB.

Conclusion: We have presented Beyond RGB, a diverse and large scale dataset that includes a large variety of illuminants and scenes captured with mobile MS and RGB cameras, together with a novel SoTA ISE algorithm as an example usage of our dataset. Our SoTA results on previous benchmarks show that adapting lessons learned from the rich research of color-constancy algorithms can lead to significant improvements on the ISE task, and that there is much untapped potential in this avenue of research now that suitable data is available. Additionally, we demonstrated the usage of our dataset by fusion of RGB and MS data from disparate sensors. We hope that Beyond RGB will enable future research into fusion of MS and RGB data, colorimetric accuracy of mobile devices, MS enhanced color-constancy, and practical applications of ISE.

References

[1] Telmo Adão, Jonáš Hruška, Luís Pádua, José Bessa, Emanuel Peres, Raul Morais, and Joaquim Joao Sousa. Hyperspectral imaging: A review on uav-based sensors, data processing and applications for agriculture and forestry. *Remote sensing*, 9(11):1110, 2017. 1

[2] Mahmoud Afifi and Michael S Brown. Deep white-balance editing. In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*, pages 1397–1406, 2020. 4

[3] J.M. Amigo. *Hyperspectral Imaging*. ISSN. Elsevier Science, 2019. 1

[4] Boaz Arad, Radu Timofte, Rony Yahel, Nimrod Morag, Amir Bernat, Yuanhao Cai, Jing Lin, Zudi Lin, Haoqian Wang, Yulun Zhang, et al. Ntire 2022 spectral recovery challenge and data set. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 863–881, 2022. 1, 2, 3

[5] Jonathan T Barron. Convolutional color constancy. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 379–387, 2015. 2, 4, 5

[6] Jonathan T Barron, Mark D Biggin, Pablo Arbelaez, David W Knowles, Soile VE Keranen, and Jitendra Malik. Volumetric semantic segmentation using pyramid context features. In *Proceedings of the IEEE international conference on computer vision*, pages 3448–3455, 2013. 5

[7] Jonathan T Barron and Yun-Ta Tsai. Fast fourier color constancy. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 886–894, 2017. 4

[8] Bryce E Bayer. Color imaging array, July 1976. 1

[9] Gershon Buchsbaum. A spatial processor model for object colour perception. *Journal of the Franklin institute*, 310(1):1–26, 1980. 4, 7

[10] Ayan Chakrabarti and Todd Zickler. Statistics of real-world hyperspectral images. In *CVPR 2011*, pages 193–200. IEEE, 2011. 1, 2, 3

[11] Chein-I Chang. *Hyperspectral data exploitation: theory and applications*. John Wiley & Sons, 2007. 1

[12] Po-Rong Chang and Tsung-Hsieh Hsieh. Constrained nonlinear optimization approaches to color-signal separation. *IEEE transactions on image processing*, 4(1):81–94, 1995. 4

[13] Dongliang Cheng, Dilip K Prasad, and Michael S Brown. Illuminant estimation for color constancy: why spatial-domain methods work and the role of the color distribution. *JOSA A*, 31(5):1049–1058, 2014. 3

[14] CIE. CIE 1931 colour-matching functions, 2 degree observer. Technical report, International Commission on Illumination, 2019. 6

[15] Mauricio Delbracio, Damien Kelly, Michael S Brown, and Peyman Milanfar. Mobile computational photography: A tour. *Annual Review of Vision Science*, 7:571–604, 2021. 1

[16] Mark S Drew and Graham D Finlayson. Analytic solution for separating spectra into illumination and surface reflectance components. *JOSA A*, 24(2):294–303, 2007. 4

[17] Graham D. Finlayson and Yuteng Zhu. Designing Color Filters That Make Cameras More Colorimetric. *IEEE Transactions on Image Processing*, 30:853–867, 2021. 1

[18] David Foster, Kinjiro Amano, and Sérgio Nascimento. Fifty hyperspectral reflectance images of outdoor scenes. 6 2022. 1, 2, 3

[19] David H Foster, Kinjiro Amano, and Sérgio MC Nascimento. Time-lapse ratios of cone excitations in natural scenes. *Vision research*, 120:45–60, 2016. 3

[20] David H Foster, Kinjiro Amano, Sérgio MC Nascimento, and Michael J Foster. Frequency of metamerism in natural scenes. *Josa a*, 23(10):2359–2372, 2006. 1, 2, 3

[21] David H Foster and Adam Reeves. Colour constancy failures expected in colourful environments. *Proceedings of the Royal Society B*, 289(1967):20212483, 2022. 3

[22] Peter Vincent Gehler, Carsten Rother, Andrew Blake, Tom Minka, and Toby Sharp. Bayesian color constancy revisited. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2008. 3

[23] Arjan Gijsenij, Theo Gevers, and Joost Van De Weijer. Computational color constancy: Survey and experiments. *IEEE transactions on image processing*, 20(9):2475–2489, 2011. 3

[24] Jian Ho, Brian V. Funt, and Mark S. Drew. Separating a color signal into illumination and surface reflectance components: Theory and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(10):966–977, 1990. 4

[25] Berthold KP Horn. Exact reproduction of colored images. *Computer Vision, Graphics, and Image Processing*, 26(2):135–167, 1984. 1

[26] Yuanming Hu, Baoyuan Wang, and Stephen Lin. Fc4: Fully convolutional color constancy with confidence-weighted pooling. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4085–4094, 2017. 4

[27] Yuanming Hu, Baoyuan Wang, and Stephen Lin. Fc4: Fully convolutional color constancy with confidence-weighted pooling. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4085–4094, 2017. 4

[28] Robert William Gainer Hunt and Michael R Pointer. *Measuring colour*. John Wiley & Sons, 2011. 1

[29] Katsushi Ikeuchi, Daisuke Miyazaki, Akifumi Ikari, Rei Kawakami, Robby T Tan, and Katsushi Ikeuchi. Separating illumination and surface spectral from multiple color signals. *Digitally archiving cultural objects*, pages 297–321, 2008. 4

[30] Herbert E Ives. The transformation of color-mixture equations from one system to another. *Journal of the Franklin Institute*, 180(6):673–701, 1915. 1

[31] Hakki Can Karaimer and Michael S Brown. A software platform for manipulating the camera imaging pipeline. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*, pages 429–444. Springer, 2016. 1

[32] Haris Ahmad Khan, Jean-Baptiste Thomas, Jon Yngve Hardeberg, and Olivier Laligant. Illuminant estimation in multispectral imaging. *JOSA A*, 34(7):1085–1098, 2017. 7

[33] Muhammad Jaleed Khan, Hamid Saeed Khan, Adeel Yousaf, Khurram Khurshid, and Asad Abbas. Modern trends in hyperspectral image analysis: A review. *IEEE Access*, 6:14118–14129, 2018. 1

[34] Sandeep Koranne and Sandeep Koranne. Hierarchical data format 5: Hdf5. *Handbook of open source tools*, pages 191–200, 2011. 4

[35] Firas Laakom, Nikolaos Passalis, Jenni Raitoharju, Jarno Nikkanen, Anastasios Tefas, Alexandros Iosifidis, and Moncef Gabbouj. Bag of color features for color constancy. *IEEE Transactions on Image Processing*, 29:7722–7734, 2020. 4

[36] Firas Laakom, Jenni Raitoharju, Jarno Nikkanen, Alexandros Iosifidis, and Moncef Gabbouj. Intel-tau: A color constancy dataset. *IEEE access*, 9:39560–39567, 2021. 5

[37] Edwin H Land and John J McCann. Lightness and retinex theory. *Josa*, 61(1):1–11, 1971. 4

[38] Jaesoong Lee, Yeonsang Park, Hyochul Kim, Young-Zoon Yoon, Woong Ko, Kideock Bae, Jeong-Yub Lee, Hyuck Choo, and Young-Geun Roh. Compact meta-spectral image sensor for mobile applications. *Nanophotonics*, 11(11):2563–2569, 2022. 1

[39] Shuwei Li, Jikai Wang, Michael S. Brown, and Robby T. Tan. Mimt: Multi-illuminant color constancy via multi-task learning, 2023. 4

[40] Yuqi Li, Qiang Fu, and Wolfgang Heidrich. Multispectral illumination estimation using deep unrolling network. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2652–2661, Montreal, QC, Canada, Oct. 2021. IEEE. 1, 2, 3, 4, 6, 7

[41] Yi-Chen Lo, Chia-Che Chang, Hsuan-Chao Chiu, Yu-Hao Huang, Chia-Ping Chen, Yu-Lin Chang, and Kevin Jou. Clcc: Contrastive learning for color constancy. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8053–8063, 2021. 4

[42] Guolan Lu and Baowei Fei. Medical hyperspectral imaging: a review. *Journal of biomedical optics*, 19(1):010901–010901, 2014. 1

[43] Rastislav Lukac and Konstantinos N Plataniotis. Color filter arrays: Design and performance analysis. *IEEE Transactions on Consumer electronics*, 51(4):1260–1267, 2005. 1

[44] Laurence T Maloney. Evaluation of linear models of surface spectral reflectance with small numbers of parameters. *JOSA A*, 3(10):1673–1683, 1986. 4

[45] Dimitris G Manolakis, Ronald B Lockwood, and Thomas W Cooley. *Hyperspectral imaging remote sensing: physics, sensors, and algorithms*. Cambridge University Press, 2016. 1

[46] Thomas Mansencal, Michael Mauderer, Michael Parsons, Nick Shaw, Kevin Wheatley, Sean Cooper, Jean D. Vandenberg, Luke Canavan, Katherine Crowson, Ofek Lev, Katrin Leinweber, Shriramana Sharma, Troy James Sobotka, Dominik Moritz, Matt Pppp, Chinmay Rane, Pavithra Eswaramoorthy, John Mertic, Ben Pearlstine, Manuel Leonhardt, Olli Niemitalo, Marek Szymanski, Maximilian Schambach, Sianyi Huang, Mike Wei, Nishant Joywardhan, Omar Wagih, Pawel Redman, Joseph Goldstone, Stephen Hill, Jedediah Smith, Frederic Savoir, Geetansh Saxena, Saransh Chopra, Ilia Sibiryakov, Tim Gates, Gajendra Pal, Nicolas Tessore, Aurélien Pierre, François-Xavier Thomas, Sabarish Srinivasan, and Tucker Downs. Colour 0.4.2, Nov. 2022. 6

[47] CS McCamy. Simulation of daylight for viewing and measuring color. *Color Research & Application*, 19(6):437–445, 1994. 4

[48] Peyman Milanfar. A tour of modern image filtering: New insights and methods, both practical and theoretical. *IEEE signal processing magazine*, 30(1):106–128, 2012. 4

[49] Junichi Nakamura. *Image sensors and signal processing for digital still cameras*. CRC press, 2017. 1

[50] Sérgio MC Nascimento, Kinjiro Amano, and David H Foster. Spatial distributions of local illumination color in natural scenes. *Vision research*, 120:39–44, 2016. 1, 2, 3

[51] Sérgio MC Nascimento, Flávio P Ferreira, and David H Foster. Statistics of spatial cone-excitation ratios in natural scenes. *JOSA A*, 19(8):1484–1490, 2002. 1, 2, 3

[52] Rang MH Nguyen, Dilip K Prasad, and Michael S Brown. Training-based spectral reconstruction from a single rgb image. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part VII 13*, pages 186–201. Springer, 2014. 1, 2, 3

[53] Edwin Olson. Apriltag: A robust and flexible visual fiducial system. In *2011 IEEE international conference on robotics and automation*, pages 3400–3407. IEEE, 2011. 5

[54] Akin Ozdemir and Kemal Polat. Deep learning applications for hyperspectral imaging: a systematic review. *Journal of the Institute of Electronics and Computer*, 2(1):39–56, 2020. 1

[55] Jussi PS Parkkinen, Jarmo Hallikainen, and Timo Jaaskelainen. Characteristic spectra of munsell colors. *JOSA A*, 6(2):318–322, 1989. 4

[56] Jonathan B Phillips and Henrik Eliasson. *Camera image quality benchmarking*. John Wiley & Sons, 2018. 3

[57] Antonio Robles-Kelly and Ran Wei. A convolutional neural network for pixelwise illuminant recovery in colour and spectral images. In *2018 24th International Conference on Pattern Recognition (ICPR)*, pages 109–114. IEEE, 2018. 1, 7

[58] Antonio Robles-Kelly and Ran Wei. A Convolutional Neural Network for Pixelwise Illuminant Recovery in Colour and Spectral Images. In *2018 24th International Conference on Pattern Recognition (ICPR)*, pages 109–114, Beijing, Aug. 2018. IEEE. 4

[59] Young-Geun Roh, Hyochul Kim, Wooshik Kim, Jaesoong Lee, Suyeon Lee, Youngho Jung, Heejin Choi, and Sangyoon Lee. Spectral image sensing in the smartphone and its applications. In *Image Sensing Technologies: Materials, Devices, Systems, and Applications X*, page PC1251408. SPIE, 2023. 1

[60] Scyllarus. Scyllarus hyperspectral data sets, 2016. 1, 2, 3

[61] Wu Shi, Chen Change Loy, and Xiaoou Tang. Deep specialized network for illuminant estimation. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14*, pages 371–387. Springer, 2016. 4

[62] Tong Su, Yu Zhou, Yao Yu, Xun Cao, and Sidan Du. Illumination separation of non-Lambertian scenes from a single hyperspectral image. *Optics Express*, 26(20):26167, Oct. 2018. 1, 3, 4

[63] Tong Su, Yu Zhou, Yao Yu, Xun Cao, and Sidan Du. Illumination separation of non-lambertian scenes from a single hyperspectral image. *Optics express*, 26(20):26167–26178, 2018. 7

[64] Jean-Baptiste Thomas. Illuminant estimation from uncalibrated multispectral images. In *2015 Colour and Visual Computing Symposium (CVCS)*, pages 1–6. IEEE, 2015. 1

[65] Fabio Tosi, Pierluigi Zama Ramirez, Matteo Poggi, Samuele Salti, Stefano Mattoccia, and Luigi Di Stefano. Rgb-multispectral matching: Dataset, learning methodology, evaluation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15958–15968, 2022. 1

[66] Joost Van De Weijer, Theo Gevers, and Arjan Gijsenij. Edge-based color constancy. *IEEE Transactions on image processing*, 16(9):2207–2214, 2007. 4, 7

[67] John Wang and Edwin Olson. Apriltag 2: Efficient and robust fiducial detection. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4193–4198. IEEE, 2016. 5

[68] Bolei Xu, Jingxin Liu, Xianxu Hou, Bozhi Liu, and Guoping Qiu. End-to-end illuminant estimation based on deep metric learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3616–3625, 2020. 4

[69] Bolei Xu, Jingxin Liu, Xianxu Hou, Bozhi Liu, and Guoping Qiu. End-to-end illuminant estimation based on deep metric learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3616–3625, 2020. 4

[70] F. Yasuma, T. Mitsunaga, D. Iso, and S.K. Nayar. Generalized Assorted Pixel Camera: Post-Capture Control of Resolution, Dynamic Range and Spectrum. Technical report, Nov 2008. 1, 2, 3, 6

[71] Yinqiang Zheng, Imari Sato, and Yoichi Sato. Illumination and reflectance spectra separation of a hyperspectral image meets low-rank matrix factorization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1779–1787, 2015. 1, 4, 7