A Multi-Hypothesis Approach to Color Constancy

Daniel Hernandez-Juarez[1], Sarah Parisot[1,2], Benjamin Busam[1,3], Aleš Leonardis[1], Gregory Slabaugh[1] and Steven McDonagh[1]

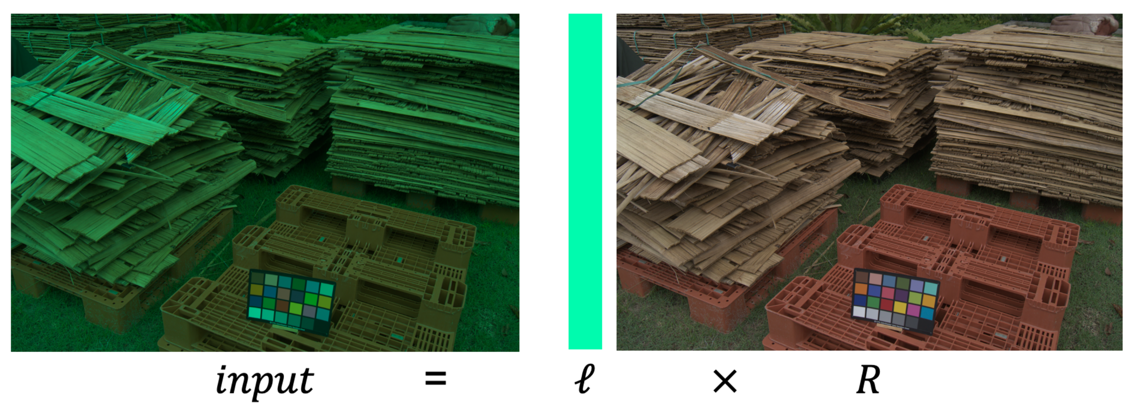
Huawei Noah's Ark Lab[1] Mila Montréal[2] Technical University of Munich[3]

dhernandez0@gmail.com, steven.mcdonagh@huawei.com

CVPR SEATTLE WASHINGTON JUNE 16-18 2020

Introduction

- **Color constancy**: the ability to account for scene light source color
- Allows extraction of color-unbiased surface information, essential for down-stream computer vision tasks



- Neural networks provide state-of-the-art performance for estimation of scene illuminant $\boldsymbol{\ell}$
- **Problem 1**: regression to illuminant point estimates do not account for multiple *plausible* solutions
- **Problem 2**: state-of-the-art learning methods are often camera-dependent yet labelled datasets are expensive to collect and typically small
- **Proposal**: a **multi-hypothesis strategy** to handle color constancy ambiguity and afford camera-agnostic training

Contributions

- Decompose the problem: **multi-hypothesis three stage** approach
- **Multi-camera learning** strategy that improves accuracy over single-camera training
- **Training-free** model adaptation to new capture devices
- **Fast** method that achieves **state-of-the-art accuracy**

Modelling Color Constancy

Assumption: the scene is illuminated by a single or dominant light source

Goal: given input image Y, estimate the global illumination color

- Let $\mathbf{y} = (y_r, y_g, y_b)$ be a pixel from image Y in linear RGB space
- Model pixel observations as the product of surface reflectance $\mathbf{r} = (r_r, r_g, r_b)$ and global illuminant $\boldsymbol{\ell} = (\ell_r, \ell_g, \ell_b)$:

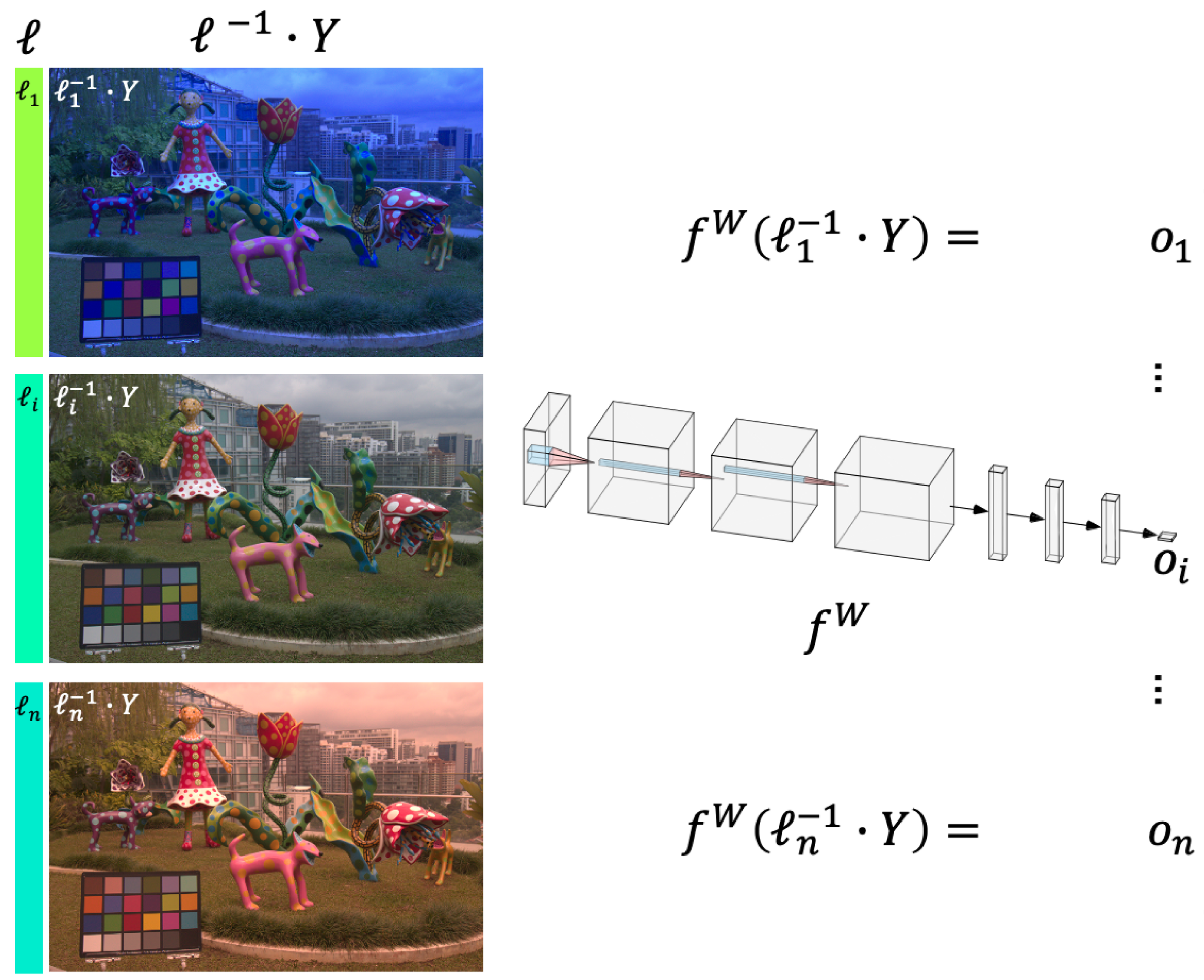
$$y_k = r_k \cdot \ell_k \qquad \text{for } k \in \{r, g, b\}.$$

- Estimate $\boldsymbol{\ell}$ in order to recover surface reflectances $R = \text{diag}(\boldsymbol{\ell})^{-1} Y$
- Enables generation of identical content yet **achromatic** appearance

But:

- **Ill-posed problem:** *infinitely many* combinations of illuminant and surface reflectance can generate an *identical* pixel observation
- Point estimates for $\boldsymbol{\ell}$ do not offer any information regarding likely alternative solutions
- Directly estimating $\boldsymbol{\ell}$ is inherently camera-specific due to camera spectral sensitivities

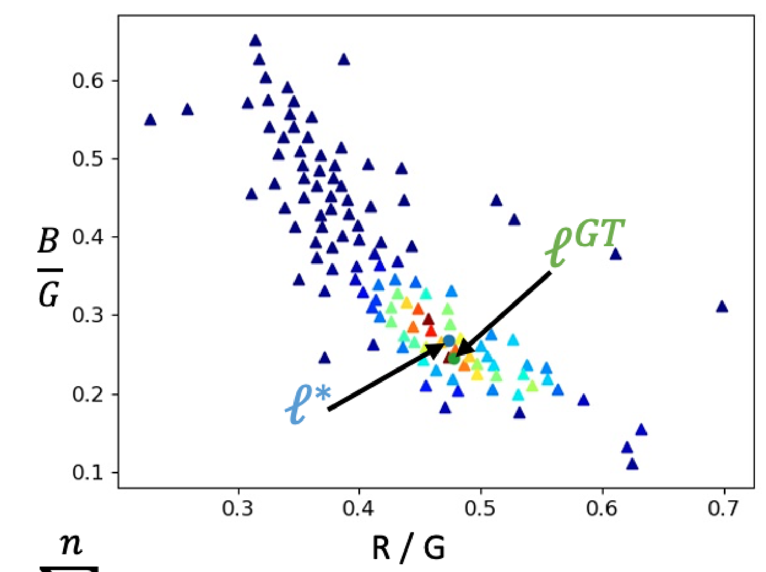
Overview



$$f^W(\boldsymbol{\ell}_1^{-1} \cdot Y) = o_1$$

$$\vdots$$

$$f^W(\boldsymbol{\ell}_n^{-1} \cdot Y) = o_n$$



$$\boldsymbol{\ell}^* = \sum_{i=1}^{n} \boldsymbol{\ell}_i \cdot softmax(G_{\boldsymbol{\ell}_i} \cdot \underbrace{f^W(\boldsymbol{\ell}_i^{-1} \cdot Y) + B_{\boldsymbol{\ell}_i}}_{o_i})$$

$$\mathcal{L} = \arccos\left(\frac{\boldsymbol{\ell}^{GT} \cdot \boldsymbol{\ell}^*}{\|\boldsymbol{\ell}^{GT}\| \|\boldsymbol{\ell}^*\|}\right)$$

Bayesian formulation

- **Idea**: account for unknown surface reflectances and scene illuminant using a probabilistic generative model
- Model the posterior distribution of illuminants $\boldsymbol{\ell}$, given the input image Y:

$$\mathrm{P}(\boldsymbol{\ell} \mid Y) = \frac{\mathrm{P}(Y \mid \boldsymbol{\ell})\,\mathrm{P}(\boldsymbol{\ell})}{\mathrm{P}(Y)}.$$

- Assuming illuminant and surface reflectance are independent allows decomposition and separate modelling of the factors
- Model the likelihood of an observed image Y, given illuminant $\boldsymbol{\ell}$ as:

$$\mathrm{P}(Y \mid \boldsymbol{\ell}) = \int_r \mathrm{P}(Y \mid \boldsymbol{\ell}, R = r)\,\mathrm{P}(R = r)\,dr = \mathrm{P}(R = \text{diag}(\boldsymbol{\ell})^{-1} Y)$$

$$\mathrm{P}(Y \mid \boldsymbol{\ell}, R = r) \text{ only non-zero for } R = \text{diag}(\boldsymbol{\ell})^{-1} Y$$

- **Intuition**: The likelihood rates whether an image, corrected under $\boldsymbol{\ell}$, *looks realistic* in terms of achromaticity
- Estimate $\boldsymbol{\ell}^*$ by optimising the quadratic cost (min. MSE Bayesian estimator)

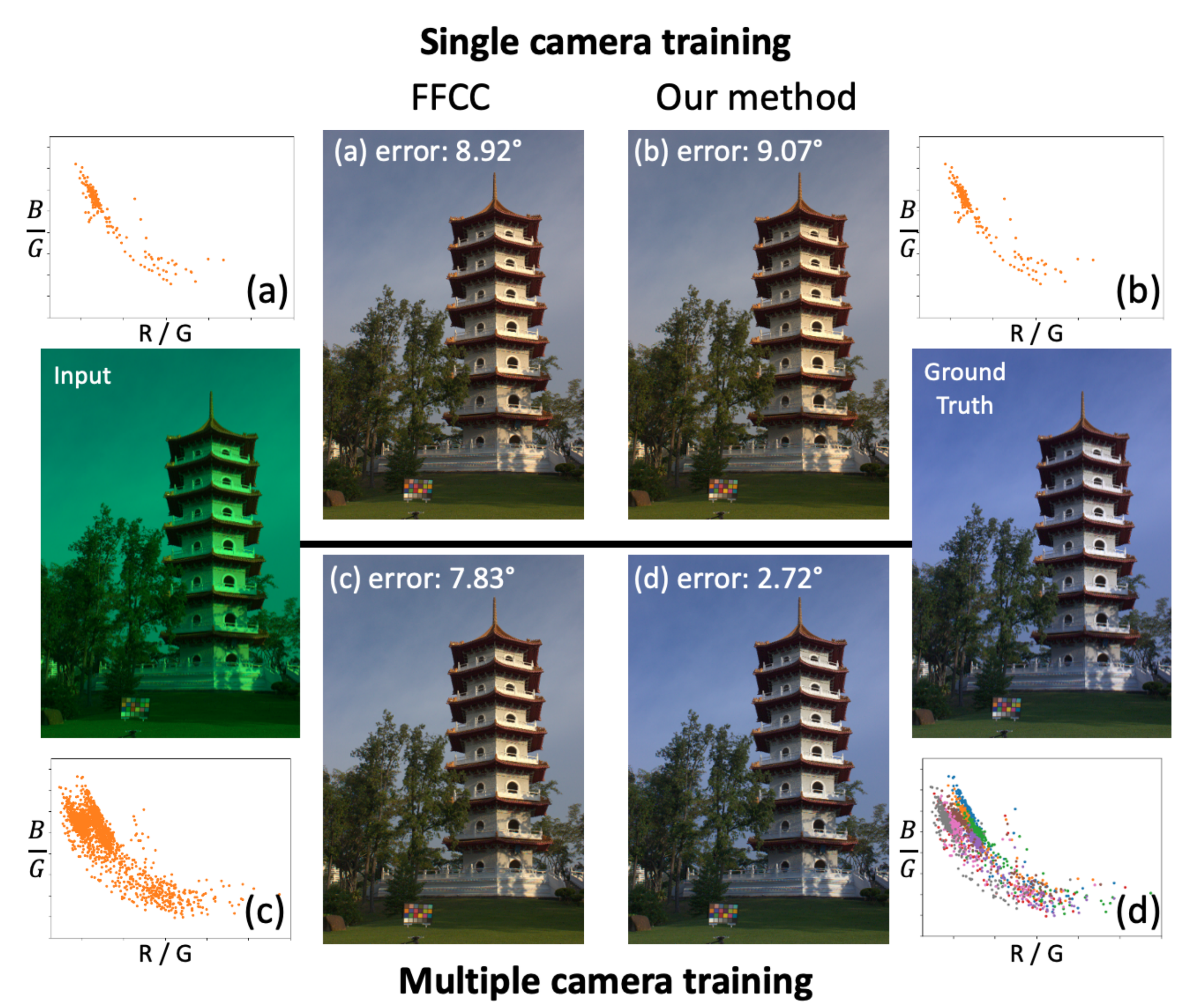
$$\boldsymbol{\ell}^* = \int_{\boldsymbol{\ell}} \boldsymbol{\ell} \cdot \mathrm{P}(\boldsymbol{\ell} \mid Y)\,d\boldsymbol{\ell}$$

Method instantiation:

1. Select a representative set of n **candidate illuminants** *e.g.* cluster training set illuminant labels $\boldsymbol{\ell}$
2. Generate n illuminant-corrected images $\boldsymbol{\ell}_i^{-1} \cdot Y$. Train model f^W to estimate the **likelihood** o_i that $\boldsymbol{\ell}_i^{-1} \cdot Y$ exhibits an achromatic light source
3. Linearly combine candidate illuminants $\boldsymbol{\ell}_i$, weighted by estimated posterior probabilities, to produce final illuminant estimate $\boldsymbol{\ell}^*$

Multi-camera training

- Model f^W provides likelihood o_i that an image is *well white balanced*: a **device-independent** learning task
- Illuminant candidates are camera-specific yet the strategy allows updating of a single set of model parameters W during training
- Typically *small* color constancy training sets can be amalgamated across capture devices towards larger data improving f^W prediction quality



Angular-Error statistics (lower better)

Method	Mean	Median	Trimean	Best 25%	Worst 25%
One model per device					
FFCC [3] (model Q)	2.37	1.50	1.69	0.46	5.76
Ours (pretrained)	2.35	1.48	1.67	0.47	5.71
Multi-device training					
FFCC [3] (model Q)	2.59	1.77	1.94	0.52	6.14
Ours (pretrained)	2.22	1.33	1.53	0.44	5.49

Table: Angular error for NUS data set using multi-device cross-validation folds

Method	Mean	Median	Trimean	Best 25%	Worst 25%
FFCC [3] (model J)	2.10	1.23	1.34	0.47	5.38
WB-sRGB [2, 1]	1.83	1.15	-	0.35	4.60
Ours	1.99	1.06	1.14	0.35	5.35

Table: Angular error for Cube challenge data set

*Please see our paper for unabridged quantitative results

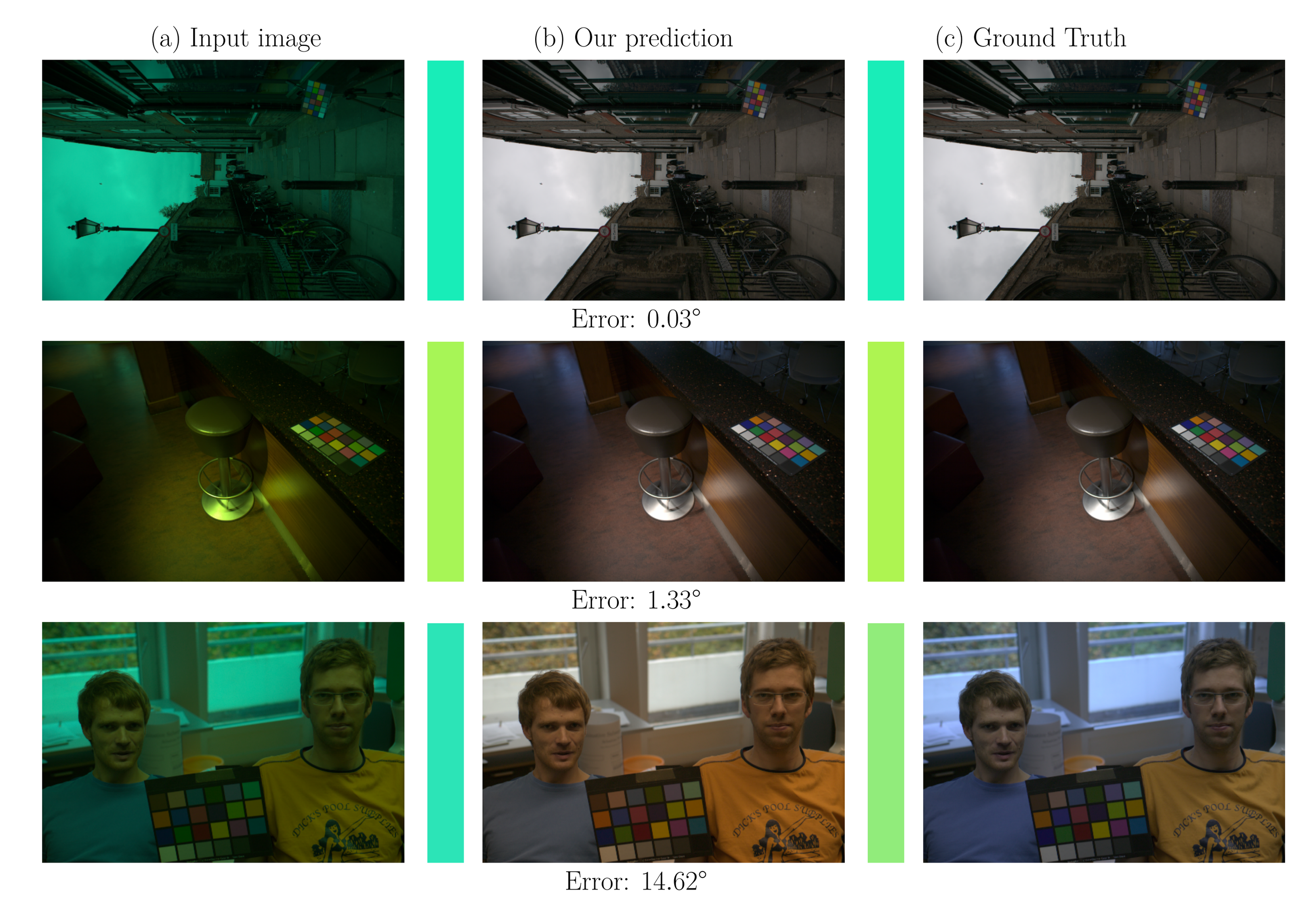
Training-free camera adaptation

- Train using amalgamated NUS and Gehler-Shi datasets
- Select candidates from Cube+ and test on separate Cube challenge datasets
- **No test camera imagery** seen during training

Method	Mean	Med.	Tri.	Best 25%	Worst 25%
Afifi *et al.* 2019 [1]	2.89	1.72	-	0.71	7.06
Ours	2.07	1.31	1.43	0.41	5.12

Table: Angular error for Cube challenge trained solely on NUS and Gehler-Shi. Candidate selection from separate Cube+ dataset

Qualitative results



Gehler-Shi test images. Results sorted by increasing angular-error and sampled uniformly to select displayed images

Links

Code: Paper:

References

[1] M. Afifi and M. Brown. Sensor-Independent Illumination Estimation for DNN Models. In *BMVC*, 2019.

[2] M. Afifi et al. When color constancy goes wrong: Correcting improperly white-balanced images. In *CVPR*, pages 1535–1544, 2019.

[3] J. T. Barron and Y. Tsai. Fast fourier color constancy. In *CVPR*, pages 6950–6958, 2017.