OPTIMAL NOISE-AWARE IMAGING WITH SWITCHABLE PREFILTERS

Zilai Gong, Masayuki Tanaka, Yusuke Monno, and Masatoshi Okutomi

Tokyo Institute of Technology, Tokyo, Japan

ABSTRACT

Most consumer digital cameras employ a single-chip image sensor with a color filter array (CFA), where the purpose of an in-camera imaging pipeline is to generate a noise-free and color-corrected standard RGB image from mosaic CFA RAW data. The joint design of camera spectral sensitivity (CSS) and the imaging pipeline has great potential to derive better imaging quality. However, since there is a trade-off between the robustness to noise and the accuracy of color reproduction, one fixed CSS cannot realize optimal imaging in terms of both aspects under various noise levels. Thus, in this paper, we propose noise-aware imaging using camera prefilters for each noise level, where we jointly design the spectral sensitivity of the prefilters, that of CFA, and imaging networks to realize optimal imaging in all noise levels. Experimental results under various noise levels demonstrate that our imaging method using the prefilters outperforms existing methods based on a fixed CSS.
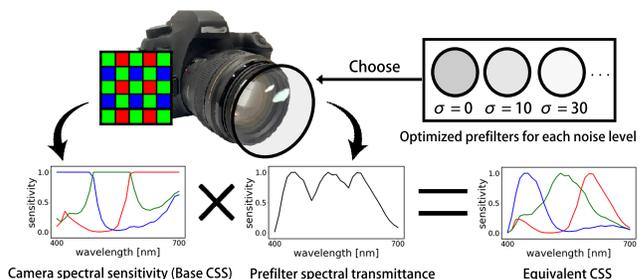
Index Terms— Camera spectral sensitivity, camera imaging pipeline, camera prefilter

1. INTRODUCTION

Most consumer digital cameras employ a single-chip image sensor with Bayer color filter array (CFA) [1]. The data captured using a CFA is called CFA RAW data, which consist of a mosaic pattern of RGB values. The CFA RAW data are processed through an in-camera imaging pipeline, which includes demosaicking [2, 3], denoising [4, 5], and color correction [6, 7], where the purpose is to generate a noise-free and color-corrected standard RGB (sRGB) [8] image.

It is well known that the design of camera spectral sensitivity (CSS) affects the color reproduction accuracy. Based on this, many studies have been conducted to design a CSS to improve colorimetric accuracy [9, 10]. As another target, the optimal design of a CSS for hyperspectral reconstruction from an RGB image has also been studied [11, 12].

Regarding the CSS design, it is also known that there is a trade-off between the robustness to noise and the accuracy of color reproduction. Empirically, a broader CSS is more robust to noise, but it brings lower color accuracy. Although some existing studies analyze this trade-off to design a CSS [13, 14], it is practically infeasible to design one fixed CSS that realizes high-quality imaging under a wide range


Fig. 1: Our imaging setup with switchable prefilters, where we jointly design the spectral transmittance of the prefilters, the camera spectral sensitivity of the CFA (Base CSS), and imaging networks associated with each prefilter.

of noise-level conditions. Thus, ideally, the CSS should be adaptive to each noise-level condition. However, it is also infeasible because the CSS is not changeable at each capturing time once it is fabricated.

In this paper, we propose a new design method for optimal noise-aware imaging using camera prefilters, which are easily switchable in front of the camera lens according to each noise-level condition, as illustrated in Fig. 1. Although the idea of designing a prefilter has already been introduced in some recent studies [15, 16], they only optimize the spectral transmittance of the prefilter based on a mathematical model such as Luther condition. In contrast, we adopt a deep learning framework and propose a new method to jointly design the spectral transmittance of the prefilter, the CSS of the CFA as a base CSS, and an imaging network to produce a final sRGB image. Experimental results under various noise levels demonstrate that our imaging method using the prefilters outperforms existing methods based on a fixed CSS.

2. PROPOSED METHOD

Figure 2 shows the overall pipeline of our method, which consists of three steps. The first step is to derive target reference CSSs for each noise level. This is performed by the joint training of a CSS and an imaging network for each given noise level. The resultant CSS of the joint training is regarded as an ideal CSS for each noise level. The second step is to jointly optimize the spectral sensitivity of the prefilters for every considered noise level and that of the CFA as a fixed base CSS, so that the differences between the resultant equiv-
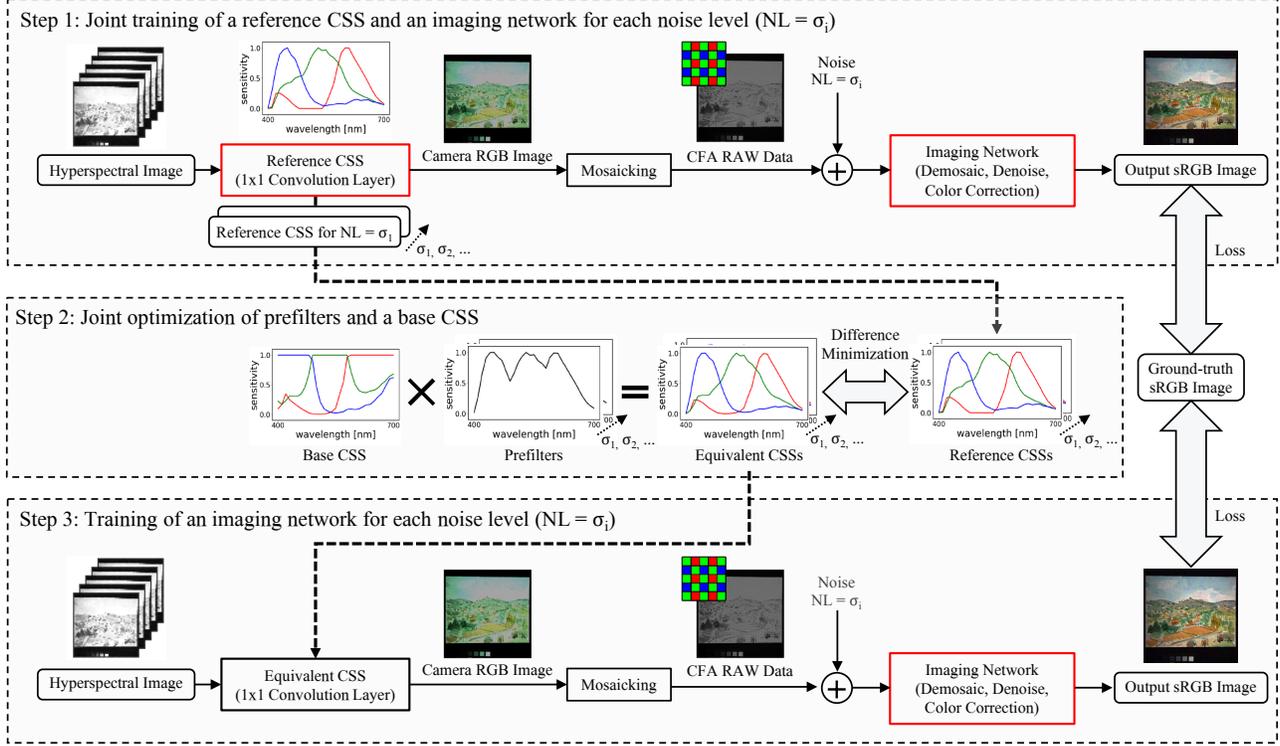
Fig. 2: The overall flow of our method for the joint design of the prefilters, the base CSS, and the imaging networks associated with each prefilter. In the first and the third steps, the parameters in the red boxes are optimized by a deep learning framework.

alent CSSs and the reference CSSs from the first step are minimized. Finally, using the equivalent CSSs from the second step, the third step performs the re-training of imaging networks associated with each equivalent CSS. We describe the details of each step below.

2.1. Reference CSSs training

It is known that an optimal CSS depends on the noise level. Also, an imaging pipeline should be optimized for the used CSS. Thus, in the first step, we jointly train a CSS and an imaging network for each noise level to derive a target reference CSS. As the results of the joint training for each noise level, we obtain reference CSSs for the set of assumed noise levels ($\sigma = 0, 1, 2, 5, 10, 20, 30$, in our experiments).

The top row of Fig. 2 shows the flow of the joint training, for which we follow the training approach of [12]. The network input is a hyperspectral image, which is first converted to a camera RGB image by a convolution layer with 1×1-sized kernel whose weights correspond to a CSS. Then, the camera RGB image is mosaicked according to the Bayer pattern to generate CFA RAW data. After adding random Gaussian noise with a considered noise level to the CFA RAW data, an imaging network is applied to convert the noisy CFA RAW data to an output sRGB image, where the imaging network performs the processes such as demosaicking, denoising, and color correction. The CSS layer and the imaging network are jointly trained in an end-to-end manner to minimize the difference between the output and the ground-truth sRGB images.

2.2. Joint optimization of prefilters and a base CSS

In the previous step, we have obtained the reference CSSs optimized for each noise level. However, as discussed in the introduction section, it is infeasible to change the CSS of the on-ship CFA at each capturing time. Thus, in the second step, we jointly optimize the prefilters and the CSS of the CFA as a fixed base CSS.

The middle row of Fig. 2 illustrates the joint optimization approach, where we minimize the differences between equivalent CSSs and reference CSSs. The equivalent CSS can be represented by the wavelength-by-wavelength product of the prefilter transmittance and the base CSS. It can be changed based on the noise level at each capturing time by switching the prefilter in front of the camera lens. The joint optimization of the prefilters and the base CSS is performed by minimizing the following cost function.

$$E = \sum_{n \in \Omega_N} \sum_{c \in \Omega_c} \sum_{\lambda \in \Omega_\lambda} \left(f_{\lambda,n} \times b_{\lambda,c} - r_{\lambda,c,n} \right)^2 \,,$$
$$s.t. \max_c b_{\lambda,c} = 1 \,, \qquad (1)$$

where λ is the wavelength, Ω_λ is a set of discretized wavelengths, c represents a color channel, Ω_c is a set of channels

$\{R, G, B\}$, n is the noise level, Ω_N is a set of noise levels to be considered, $f_{\lambda,n}$ is the prefilter transmittance of the wavelength λ for noise level n, $b_{\lambda,c}$ is the base CSS of the wavelength λ for c-channel, and $r_{\lambda,c,n}$ is the reference CSS of the wavelength λ for c-channel and for noise level n.

The minimization of Eq. (1) can be performed by an alternative optimization, where we apply an alternative process to find the optimal prefilters for the given base CSS and reference CSSs and then find the optimal base CSS for the given prefilters and reference CSSs. In the optimization, we put the condition for the max sensitivity of the base CSS to maximize the signal-to-noise ratio.

2.3. Imaging network training

In the previous step, we have designed the prefilters and the base CSS so that we can adaptively change the equivalent CSSs. Since the equivalent CSSs obtained in the previous step are not exactly the same as the reference CSSs, we retrain the imaging networks for each equivalent CSS in the last step. The bottom row of Fig. 2 shows the training process. Although the overall pipeline is the same as the first step of Sec. 2.1, we here fix the CSS layer as the equivalent CSS and only train the imaging network.

In the application phase, we select the prefilter according to the noise level and apply the imaging network trained using the associated equivalent CSS.

3. EXPERIMENTAL RESULTS

3.1. Datasets and training details

We evaluated the proposed method with two hyperspectral datasets of CAVE [17] and TokyoTech [18]. We have developed the prefilters, the base CSS, and the imaging networks for each dataset because those datasets have different wavelength ranges. The CAVE and the TokyoTech datasets include 32 and 30 scenes, respectively. We used eight scenes of each dataset for the testing, and the rest scenes for the training. For the imaging network, we used the U-net architecture [19][1]. We assumed CIE D65 as illumination and applied a zero-mean Gaussian noise model as in [13].

We trained our networks with 128×128-sized image patches while we evaluated for the whole image in the test phase. In the training batch generation, we randomly sampled four different images from the training data. Then, we generated 32 total image patches by cropping eight image patches from each sampled image. For the data augmentation, we randomly applied three geometrical transformations: horizontal flip, vertical flip, and transpose.

---

[1]Our code will be available on our project page: `http://www.ok.sc.e.titech.ac.jp/res/OptimalFilter/index.html`
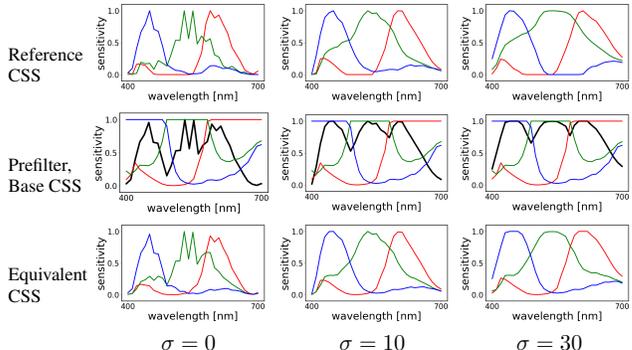
Fig. 3: Examples of the reference CSSs, the prefilters, the base CSS, and the equivalent CSSs obtained by our method for the CAVE dataset, where red, green, and blue lines represent the color CSSs and the black line represents the prefilters.

3.2. Designed prefilters and a base CSS

The first row of Fig. 3 shows the reference CSSs, which are obtained by the joint training process of our first step using the CAVE dataset. We can observe that the joint training process makes the shape of the reference CSS wider for a larger noise level, which is reasonable in the sense of signal-to-noise ratio. The second row in Fig. 3 shows the prefilters and the base CSS derived by our second step, where red, green, and blue lines represent the base CSS and the black line represents the prefilters. The third row in Fig. 3 shows the equivalent CSSs, which are the product of the prefilters and the base CSS. From the results, we can confirm that the shapes of the equivalent CSSs are sufficiently close to the reference CSSs.

3.3. Comparisons with other methods based on fixed CSSs

We compared the proposed method with other methods based on existing CSSs. As for the existing CSSs for comparison, we selected the CSSs of Sony NEX-5N, Canon 300D, and Olympus E-PL2 from the CSS database of [20]. For the imaging process of the existing CSSs, we applied signal-processing-based and network-based imaging methods. The signal-processing-based imaging (SP) consists of the sequential applications of residual interpolation [3] for demosaicking, block-matching and 3D filtering (BM3D) [4] for denoising, and standard linear color correction [6]. For the network-based imaging (Net), we applied the same U-net architecture [19] as the proposed method and trained the imaging networks for each existing CSS.

We visually compare the output sRGB images in Fig. 4 and Fig. 5. In both figures, the top row shows the output images and the bottom row shows the RMSE maps compared with the ground-truth images. The darker the color of the RMSE map, the lower the error between the output and ground-truth images. To provide a visual reference for the strength of the noise, we included the Noisy case in the comparison, which is the resulting image of applying residual
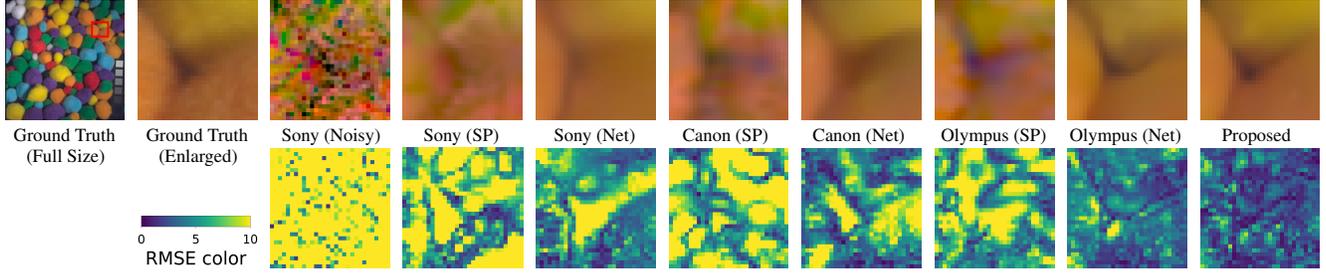
Fig. 4: Visual comparisons and RMSE maps of CAVE dataset for noise level $\sigma = 30$, where Noisy, SP, and Net represent signal processing imaging without denoising, signal processing imaging, and network-based imaging, respectively.
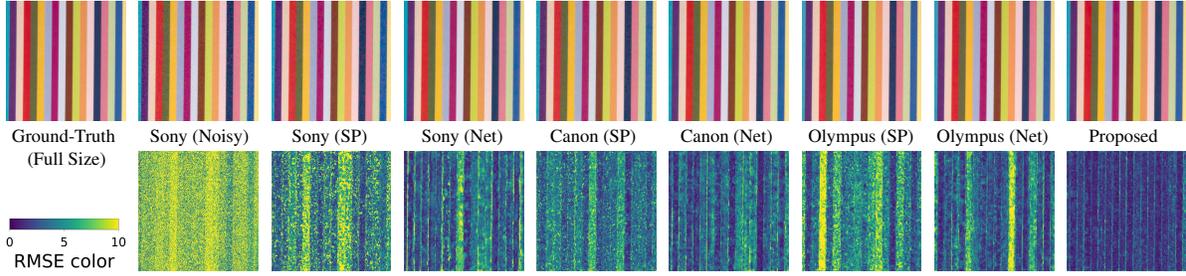
Fig. 5: Visual comparisons and RMSE maps of TokyoTech dataset for noise level $\sigma = 10$, where Noisy, SP, and Net represent signal processing imaging without denoising, signal processing imaging, and network-based imaging, respectively.

Table 1: Numerical comparisons of CAVE dataset, where Noisy, SP, and Net represent signal processing imaging without denoising, signal processing imaging, and network-based imaging, respectively.

Metric	Noise level	Sony			Canon		Olympus		Ours
		Noisy	SP	Net	SP	Net	SP	Net	
CPSNR	0	43.16	43.16	46.37	43.19	47.02	41.51	45.61	**48.07**
	1	41.61	42.20	45.46	42.61	45.82	41.28	44.88	**46.94**
	2	39.29	40.95	44.17	41.81	44.70	40.90	44.12	**46.08**
	5	33.95	37.77	41.63	39.46	42.35	39.59	42.51	**43.59**
	10	28.70	34.03	39.11	36.32	40.35	37.57	40.52	**41.17**
	20	23.07	29.52	36.50	32.21	37.52	34.29	37.99	**38.54**
	30	19.81	26.66	34.39	29.48	35.73	31.82	36.55	**36.94**
	Avg.	32.80	36.33	41.09	37.87	41.93	38.14	41.74	**43.05**
SSIM	0	0.986	0.986	0.995	0.987	0.996	0.982	0.994	**0.997**
	1	0.981	0.983	0.994	0.985	0.995	0.981	0.994	**0.997**
	2	0.970	0.978	0.993	0.982	0.993	0.980	0.990	**0.996**
	5	0.901	0.958	0.987	0.971	0.989	0.974	0.988	**0.992**
	10	0.783	0.914	0.980	0.943	0.982	0.958	0.983	**0.988**
	20	0.578	0.830	0.966	0.877	0.972	0.908	0.973	**0.979**
	30	0.435	0.762	0.953	0.814	0.957	0.850	0.959	**0.972**
	Avg.	0.805	0.916	0.981	0.937	0.983	0.948	0.983	**0.989**

Table 2: Numerical comparisons of TokyoTech dataset, where Noisy, SP, and Net represent signal processing imaging without denoising, signal processing imaging, and network-based imaging, respectively.

Metric	Noise level	Sony			Canon		Olympus		Ours
		Noisy	SP	Net	SP	Net	SP	Net	
CPSNR	0	37.75	37.75	42.60	38.74	42.86	35.44	40.44	**44.35**
	1	37.11	37.02	41.10	38.21	42.07	35.30	39.91	**43.69**
	2	35.84	35.97	39.87	37.38	41.03	35.05	39.09	**42.41**
	5	32.01	33.58	36.84	35.33	38.46	34.20	37.57	**39.72**
	10	27.62	31.07	34.79	33.05	35.98	32.91	35.69	**37.54**
	20	22.47	27.93	32.03	30.17	33.45	30.98	33.63	**34.56**
	30	19.34	25.73	30.78	28.18	31.79	29.47	32.30	**33.22**
	Avg.	30.31	32.72	36.86	34.44	37.95	33.34	36.95	**39.36**
SSIM	0	0.972	0.972	0.992	0.971	0.991	0.967	0.989	**0.995**
	1	0.969	0.969	0.988	0.970	0.989	0.966	0.987	**0.994**
	2	0.960	0.963	0.983	0.966	0.987	0.964	0.986	**0.992**
	5	0.908	0.942	0.970	0.954	0.977	0.956	0.978	**0.986**
	10	0.795	0.906	0.953	0.930	0.965	0.940	0.966	**0.974**
	20	0.600	0.844	0.926	0.883	0.945	0.905	0.949	**0.958**
	30	0.462	0.791	0.910	0.839	0.925	0.868	0.934	**0.945**
	Avg.	0.809	0.912	0.960	0.930	0.968	0.938	0.970	**0.978**

interpolation for demosaicking and linear color correction, but without applying any denoising method. From the visual comparison, we can confirm that the proposed method effectively suppresses the other methods at both $\sigma = 30$ and $\sigma = 10$. Additional comparisons can be seen in a supplementary material available on our project page (see footnote 1).

Tables 1 and 2 summarize the numerical comparisons for CAVE and TokyoTech datasets in color peak signal-to-noise ratio (CPSNR) and structural similarity index (SSIM) [21]. Both metrics represent higher is better. Those comparisons demonstrate that our proposed method outperforms existing methods in all the noise levels for both datasets.

4. CONCLUSION

In this paper, we have proposed a novel method for noise-aware imaging using switchable prefilters. To derive optimal imaging based on a deep learning framework, we have proposed a method for jointly designing the spectral transmittance of the prefilters, the CSS of the CFA, and the imaging networks associated with each prefilter. We have experimentally validated that the proposed imaging method using the prefilters outperforms the existing imaging methods based on a fixed CSS for a wide range of noise levels.

5. REFERENCES

[1] B. Bayer, "Color imaging array," *U.S. Patent 3971065*, 1976.

[2] K. Hirakawa and T. W. Parks, "Adaptive homogeneity-directed demosaicking algorithm," *IEEE Trans. on Image Processing*, vol. 14, no. 3, pp. 360–369, 2005.

[3] D. Kiku, Y. Monno, M. Tanaka, and M. Okutomi, "Beyond color difference: Residual interpolation for color image demosaicking," *IEEE Trans. on Image Processing*, vol. 25, no. 3, pp. 1288–1300, 2016.

[4] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. on Image Processing*, vol. 16, no. 8, pp. 2080–2095, 2007.

[5] H. Akiyama, M. Tanaka, and M. Okutomi, "Pseudo four-channel image denoising for noisy CFA raw data," *Proc. of IEEE Int. Conf. on Image Processing (ICIP)*, pp. 4778–4782, 2015.

[6] P. M. Hubel, J. Holm, G. D. Finlayson, and M. S. Drew, "Matrix calculations for digital photography," *Proc. of Color and Imaging Conference (CIC)*, pp. 105–111, 1997.

[7] R. Yamakabe, Y. Monno, M. Tanaka, and M. Okutomi, "Tunable color correction for noisy images," *Journal of Electronic Imaging*, vol. 29, no. 3, pp. 033012–1–24, 2020.

[8] M. Anderson, R. Motta, S. Chandrasekar, and M. Stokes, "Proposal for a standard default color space for the internet – sRGB," *Proc. of Color and Imaging Conference (CIC)*, pp. 238–245, 1996.

[9] G. Sharma and H. J. Trussell, "Figures of merit for color scanners," *IEEE Trans. on Image Processing*, vol. 6, no. 7, pp. 990–1001, 1997.

[10] G. D. Finlayson and Y. Zhu, "Designing color filters that make cameras more colorimetric," *IEEE Trans. on Image Processing*, vol. 30, pp. 853–867, 2021.

[11] B. Arad and O. Ben-Shahar, "Filter selection for hyperspectral estimation," *Proc. of IEEE Int. Conf. on Computer Vision (ICCV)*, pp. 3153–3161, 2017.

[12] S. Nie, L. Gu, Y. Zheng, A. Lam, N. Ono, and I. Sato, "Deeply learned filter response functions for hyperspectral reconstruction," *Proc. of IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 4767–4776, 2018.

[13] M. J. Vrhel and H. J. Trussell, "Optimal color filters in the presence of noise," *IEEE Trans. on Image Processing*, vol. 4, no. 6, pp. 814–823, 1995.

[14] N. Shimano, "Optimization of spectral sensitivities with gaussian distribution functions for a color image acquisition device in the presence of noise," *Optical Engineering*, vol. 45, no. 1, pp. 01320–1–8, 2006.

[15] M. Vrhel, "Improved camera color accuracy in the presence of noise with a color prefilter," *Proc. of Color and Imaging Conference (CIC)*, pp. 187–192, 2020.

[16] Y. Zhu and G. D Finlayson, "Designing a color filter with high overall transmittance for improving the color accuracy of digital cameras," *Proc. of Color and Imaging Conference (CIC)*, pp. No.29–1–6, 2021.

[17] F. Yasuma, T. Mitsunaga, D. Iso, and S. K. Nayar, "Generalized assorted pixel camera: Postcapture control of resolution, dynamic range, and spectrum," *IEEE Trans. on Image Processing*, vol. 19, no. 9, pp. 2241–2253, 2010.

[18] Y. Monno, S. Kikuchi, M. Tanaka, and M. Okutomi, "A practical one-shot multispectral imaging system using a single image sensor," *IEEE Trans. on Image Processing*, vol. 24, no. 10, pp. 3048–3059, 2015.

[19] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Proc. of Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 234–241, 2015.

[20] J. Jiang, D. Liu, J. Gu, and S. Süsstrunk, "What is the space of spectral sensitivity functions for digital color cameras?," *Proc. of Workshop on Applications of Computer Vision (WACV)*, pp. 168–179, 2013.

[21] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.