Supplementary Material: "Spectral Sensitivity Estimation with an Uncalibrated Diffraction Grating"

Lilika Makabe[1] Hiroaki Santo[1] Fumio Okura[1] Michael S. Brown[2] Yasuyuki Matsushita[1,3]
[1]The University of Osaka [2]York University [3]Microsoft Research Asia – Tokyo
{makabe.lilika,santo.hiroaki,okura,yasumat}@ist.osaka-u.ac.jp mbrown@yorku.ca

In this supplementary material,

A. we provide a detailed description of the experimental setup.
B. we present additional experimental results using different light sources and diffraction grating sheets.
C. we present additional experimental results of pixel-to-wavelength mapping of our method.
D. we provide a further derivation of the pixel-to-wavelength mapping representation.
E. we provide a detailed description of the basis representations for grating efficiency.
F. we provide a detailed description of the linear system in the proposed solution method.
G. we provide an additional comparison with the existing diffraction grating-based method.
H. we present a detailed conditions for one-to-one pixel-to-wavelength mapping to be hold.
I. we evaluate the consistency of grating efficiency estimations across different cameras.
J. we present the raw measurements captured in the real-world experiments.

A. Real-world experimental details

Figure S1 (Left) shows the our experimental setup with approximate distances annotated. The incoming light goes through the 3D-printed slit with the $2.5\,\mathrm{mm}$ width.

The images are captured in a completely dark room setting with the RAW format, with all camera parameters manually fixed, including white balance, ISO, and exposure time. The white balance is set to daylight and the ISO to 100 throughout the experiment.

B. Experiment results using different light sources and grating sheets

To verify the effects of different incoming light spectra and differences among diffraction grating sheets, we compare the results using three different light sources and three different diffraction grating sheets. For the light sources, we use the following three LEDs:

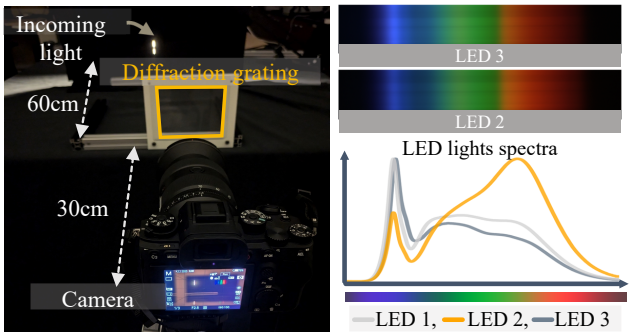

Figure S1. The left-hand side image shows our experimental setup with the specific distance between the light source with a slit, the diffraction grating, and the camera. The second row shows the plots of the light spectra of LED 1, LED 2, and LED 3 and examples of diffracted observations. x-axis and y-axis correspond to wavelength and the intensity in each wavelength, respectively. The observations are adjusted for visualization.

LED 1: "Godox LITEMONS LED6R (GODOX Photo Equipment Co., Ltd.)," with a color temperature of $6500\,\mathrm{K}$.
LED 2: The same product as LED 1, with a color temperature of $3200\,\mathrm{K}$.
LED 3: "Phone Selfie Light (Mionondi)," with a color temperature of $3200\,\mathrm{K}$.

The light spectra and the examples of the diffracted observations are shown in Fig. S1. For the diffraction grating sheets, we use the following three sheets:

Grating 1: $500\,\mathrm{slits/mm}$, "500 lines/mm Linear Diffraction Grating Sheet" sold by Bartovation.
Grating 2: $1000\,\mathrm{slits/mm}$, "1000 lines/mm Linear Diffraction Grating Sheet" sold by Bartovation.
Grating 3: $500\,\mathrm{slits/mm}$, "Diffraction Grating Sheet Replica 500" sold by Kenis.

We use LED 1 and Grating 1 in the main paper.

Figure S2 shows the results for Sony $\alpha 1$ using LED 1, LED 2, and LED 3 with Grating 1. All results demonstrate nearly identical camera spectral sensitivities, indicat-

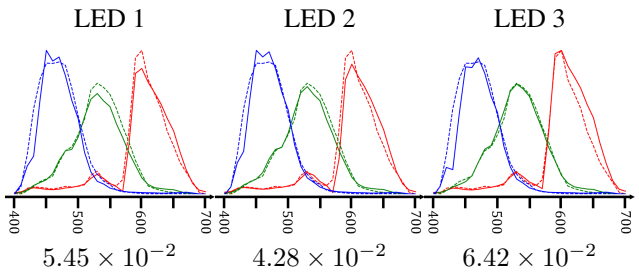

Figure S2. Estimated results of our method using different light sources for Sony α1 with Grating 1

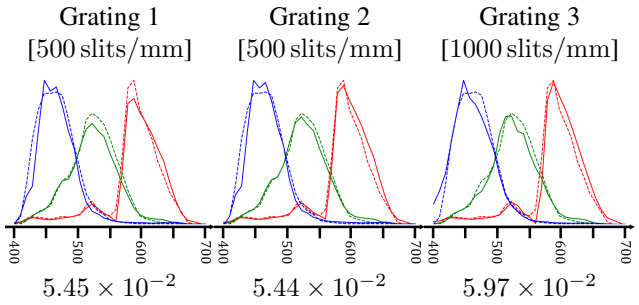

Figure S3. Estimated results of our method using different diffraction grating sheets for Sony α1 with LED 1

ing the stability of our method across different light sources. Figure S3 compares the use of different diffraction grating sheets on Sony α1 using LED 1. Overall, the estimated results remain consistent regardless of the diffraction grating choice.

C. Experiment results of pixel-to-wavelength mapping

To evaluate mapping estimation using a light source with spiky and non-spiky spectrum, we show the qualitative results of the pixel-to-wavelength mapping by our methods in Fig. S4, with quantitative results shown in Table S1. We used the following relative error metric, similar to the main paper:

$$\mathrm{RE} = \frac{1}{\lambda_M - \lambda_m} \sqrt{\frac{\|\boldsymbol{\lambda}^* - \boldsymbol{\lambda}\|_2^2}{N}}, \tag{1}$$

where $\boldsymbol{\lambda} = [g(p_1), \ldots, g(p_n)] \in \mathbb{R}^n$ is the corresponding wavelength interpolated by pixel-to-wavelength mapping function $g(p)$, and $\boldsymbol{\lambda}$ and $\boldsymbol{\lambda}^*$ is the interpolated wavelength by ground-truth and estimated pixel-to-wavelength mapping function, respectively.

The experiment shows that the pixel-to-wavelength mapping using fluorescent light can achieve around 0.3% error regardless of the target camera sensitivity. While the mapping estimation using an LED has a larger error compared to

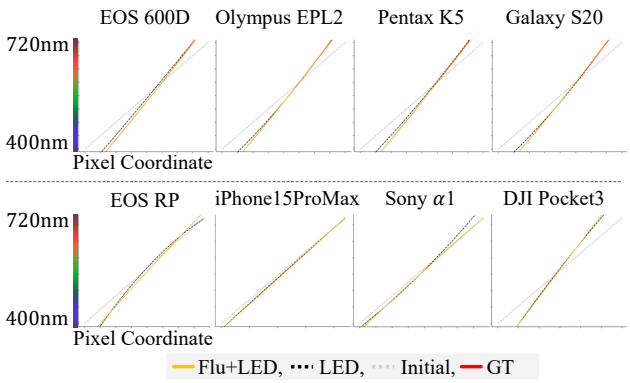

Figure S4. Plots of pixel-to-wavelength mapping from the synthetic experiment in the main paper (top row) and the real-world experiment (second row.) The legend is shown in the figure. We put the ground-truth plot only for the synthetic experiment, since we do not have the ground-truth mapping in the real-world setting. The Initial corresponds to the initial guess of the pixel-to-wavelength mapping for Ours (LED), where $a = 0$, $b = f/n$, $c = 0$, as described in the main paper. x-axis corresponds to the pixel coordinate, and y-axis to wavelength. Estimation using a non-spiky spectrum can achieve comparable results to estimation using a spiky spectrum.

	EOS 600D	Olympus EPL2	Pentax K5	Sumsung Galaxy S20
Flu+LED	0.31	0.31	0.31	0.31
LED	2.18	2.13	2.34	2.10

Table S1. Synthetic quantitative results of pixel-to-wavelength mapping. "Fluorescent" corresponds to the estimation using a light source with a spiky spectrum, and "LED" to the estimation with a non-spiky spectrum. The values are scaled by 10^2 as percentage for display purposes.

the fluorescent one, and the error is more sensitive to variations in camera spectral sensitivity due to the relatively low-frequency light spectrum, the results still show at most 2% error. This yields comparable results to the estimation using fluorescent light, even with a rough initial guess, in both the synthetic and real-world experiments.

D. Pixel-to-wavelength mapping representation

In this section, we provide a rationale for approximating the pixel-to-wavelength mapping $\lambda(p)$ using a quadratic function.

Assuming an ideal diffraction grating, there is a linear relation between a certain wavelength λ and the sin value of k^{th} diffraction angle $\theta_k(\lambda)$ as

$$d \sin \theta_k(\lambda) = k\lambda, \tag{2}$$

where d is the slit width of the diffraction grating, and the

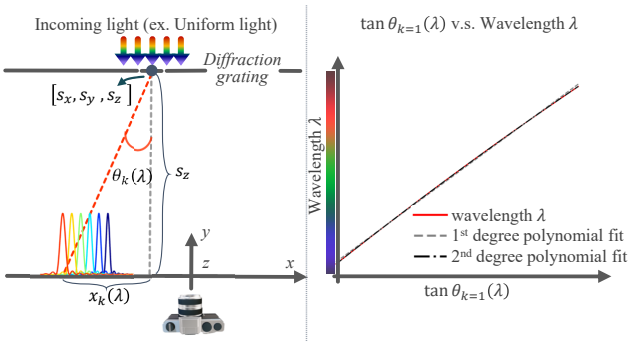

Fig. (a) Fig. (b)

Figure S5. (a) Visual illustration of the geometric relation of the diffraction grating and the diffraction angle. (b) The plot of λ with respect to $\tan(\theta_{k=1})$. In the visible wavelength domain, 1^{st} or2^{nd} degree of polynomial fit can sufficiently represent the original function.

diffraction angle $\theta_k(\lambda)$ describes the k^{th} intensity maxima position $x_k(\lambda)$ as illustrated in Fig. S5 (a).

With $500\,\text{slits/mm}$ diffraction grating, which is used in our real-world experiments, for example, the slit width d is $d = \frac{1}{500}\text{mm} = 2.0 \times 10^{-6}\,\text{m}$. Given the visible wavelength domain, which is approximately in $[380\,\text{nm}, 720\,\text{nm}]$, we can compute the possible range of sine of the 1^{st}diffraction grating angle as:

$$\begin{aligned} \frac{3.8 \times 10^{-7}}{2 \times 10^{-6}} \leq \sin\theta_{k=1} \leq \frac{7.2 \times 10^{-7}}{2 \times 10^{-6}} \\ \therefore 0.190 \leq \sin\theta_{k=1} \leq 0.360 \\ \therefore 0.191 \leq \theta_{k=1}(\lambda) \leq 0.369. \end{aligned} \tag{3}$$

Figure S5 (b) shows the plot of $y = \lambda$ with respect to $\tan\theta_k(\lambda)$ where $\theta_{k=1}(\lambda) \in [0.191, 0.369]$, accompanied with the 1^{st} and 2^{nd} degree of polynomial fit results of $y = \lambda$. It is obvious that the wavelength λ and $\tan\theta_k(\lambda)$ is almost in the affine relation in the given domain, and at most 2^{nd} degree of polynomial approximation is good enough to represent the wavelength λ using a tangent value of diffraction angle, $\tan\theta_k(\lambda)$.

At the same time, $\tan\theta_k(\lambda)$ can be described as

$$\tan\theta_k(\lambda) = \frac{x_k(\lambda)}{s_z}, \tag{4}$$

where $\begin{bmatrix} s_x, & s_y, & s_z \end{bmatrix}^{\mathsf{T}} \in \mathbb{R}^3$ is the center coordinate of the diffraction grating in the camera coordinate system. Thus similarly, λ can be represented using 2^{nd} degree of polynomial of $x_k(\lambda)$.

Furthermore, assuming the image plane and the diffraction grating sheet are parallel with each other, then we can define the intensity maxima position in the camera coordinate system $\mathbf{x}_k(\lambda) \in \mathbb{R}^3$ as

$$\mathbf{x}_k(\lambda) = \begin{bmatrix} x_k(\lambda) + s_x \\ s_y \\ s_z, \end{bmatrix}. \tag{5}$$

Assuming a pin-hole camera model, the corresponding pixel coordinate along x-axis, or along the diffraction direction, $p_k(\lambda)$ can be represented in an affine relation as

$$p_k(\lambda) = f_x \frac{x_k(\lambda) + s_x}{s_z} + c_x, \tag{6}$$

where f_x and $c_x \in \mathbb{R}_+$ represent focal length and the image center in the horizontal direction, respectively.

Overall, since λ can well represented by 2^{nd} degree of polynomial of $\tan\theta_k(\lambda)$ and since $\tan\theta_k(\lambda)$ and the corresponding pixel position $p_k(\lambda)$ is in an affine relation to $\tan\theta_k(\lambda)$, wavelength λ can be represented by quadratic function of the pixel coordinate $p(\lambda)$, and pixel-to-wavelength mapping $\lambda = g(p)$ is represented as Eq. (13) in the main paper.

E. Basis representations for grating efficiency

We define a Fourier basis for the grating efficiency function $\mathbf{B}_\eta = [\mathbf{b}_k]_{k=1,\ldots,t} \in \mathbb{R}^{f \times t}$ as follows:

$$\mathbf{b}_k = \begin{cases} \cos(2\pi k f_i) & \text{if } k \text{ is odd,} \\ \sin(2\pi k f_i) & \text{if } k \text{ is even and } k \neq 0, \\ 1 & \text{if } k = 0, \end{cases} \tag{7}$$

where t denotes a number of basis, λ_l, λ_u are the lower and upper bound of the wavelength range, respectively, and $f_i = \frac{\lambda_i}{\lambda_u - \lambda_l}$.

F. Estimation details

Here we detail the vectors and matrices that appear in Eq. (12) of the main paper:

$$\mathbf{x}^* = \underset{\mathbf{x}}{\text{argmin}} \|\mathbf{A}_{\text{dif}}\mathbf{x}\|_2^2 \;\text{ s.t. }\; \begin{bmatrix} \mathbf{0} & \mathbf{A}_{\text{dir}} \end{bmatrix} \mathbf{x} = \mathbf{m}_{\text{dir}}. \tag{8}$$

In the constraint, $\mathbf{A}_{\text{dir}}$ is a $3 \times 3b_s$ dimensional matrix written as

$$\mathbf{A}_{\text{dir}} = \begin{bmatrix} \mathbf{e}^\top \mathbf{B}_{s(R)} & 0 & 0 \\ 0 & \mathbf{e}^\top \mathbf{B}_{s(G)} & 0 \\ 0 & 0 & \mathbf{e}^\top \mathbf{B}_{s(B)} \end{bmatrix},$$

where $\mathbf{B}_{s(R)}$, $\mathbf{B}_{s(G)}$, and $\mathbf{B}_{s(B)}$ are basis matrices for each RGB color channel. The accompanying zero matrix $\mathbf{0}$ is $3 \times b_\eta$ dimensional.

In the objective, $\mathbf{A}_{\text{dif}}$ becomes a $3f \times (b_\eta + 3b_s)$ dimensional matrix as

$$\mathbf{A}_{\text{dif}} = \begin{bmatrix} \text{diag}\left(\mathbf{a}_{(R)}\right)\mathbf{B}_\eta & -\mathbf{B}_{s(R)} & 0 & 0 \\ \text{diag}\left(\mathbf{a}_{(G)}\right)\mathbf{B}_\eta & 0 & -\mathbf{B}_{s(G)} & 0 \\ \text{diag}\left(\mathbf{a}_{(B)}\right)\mathbf{B}_\eta & 0 & 0 & -\mathbf{B}_{s(B)} \end{bmatrix},$$

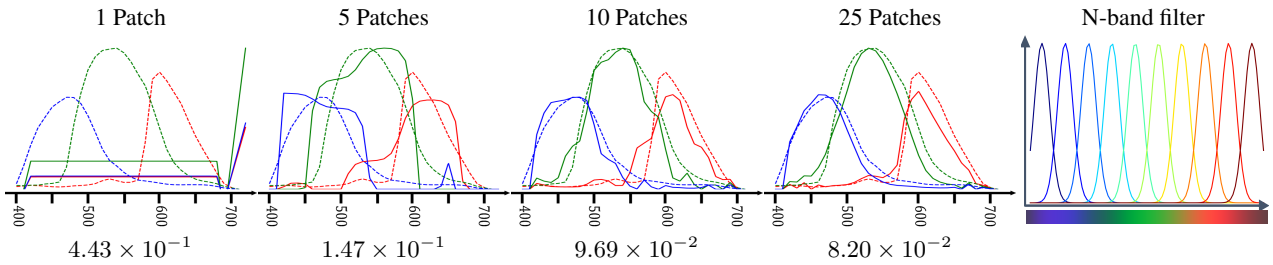

Figure S6. Estimation results of the existing diffraction grating-based method [1] using different numbers of patches in the transmissive color chart. Since there is no known dataset for transmissive filters, we synthesized N-band transmissive filters. We implemented them using shifted Gaussian distributions, ensuring that the visible wavelength range was always fully covered by the N filters. The rightmost figure shows an example filter spectrum for the $N = 10$ case.

where vector $\mathbf{a}_{(c)}$ is defined as

$$\mathbf{a}_{(c)} = \mathrm{diag}\left(\mathbf{e}^{-1}\right) \mathbf{W}^{\dagger} \mathbf{m}_{\mathrm{dif}(c)}$$

for color channel $c \in \{R, G, B\}$, and $\mathbf{m}_{\mathrm{dif}(c)}$ is a vector of diffracted light observations in color channel c. $\mathbf{x} = \begin{bmatrix} \mathbf{c}_{\eta}^{\mathsf{T}}, & \mathbf{c}_{s(R)}^{\mathsf{T}}, & \mathbf{c}_{s(G)}^{\mathsf{T}}, & \mathbf{c}_{s(B)}^{\mathsf{T}} \end{bmatrix}^{\mathsf{T}} \in \mathbb{R}^{b_\eta + 3b_s}$ is the vector of unknowns in total.

To solve this problem, we introduce Lagrange multiplier $\boldsymbol{\mu} \in \mathbb{R}^3$ whose dimension corresponds to the number of equations in the linear constraints from the direct light observation $\mathbf{A}_{\mathrm{dir}}$. With the Lagrange multiplier $\boldsymbol{\mu}$, the Lagrangian $\mathcal{L}$ is defined as:

$$\mathcal{L} = \|\mathbf{A}_{\mathrm{dif}}\mathbf{x}\|_2^2 + \boldsymbol{\mu}^{\mathsf{T}} \left(\begin{bmatrix} \mathbf{0} & \mathbf{A}_{\mathrm{dir}} \end{bmatrix} \mathbf{x} - \mathbf{m}_{\mathrm{dir}} \right),$$

whose stationary point should satisfy $\frac{\partial \mathcal{L}}{\partial \mathbf{x}} = \mathbf{0}$ and $\frac{\partial \mathcal{L}}{\partial \boldsymbol{\mu}} = \mathbf{0}$. It results in the following linear system of equations:

$$\mathbf{A} \begin{bmatrix} \mathbf{x} \\ \boldsymbol{\lambda} \end{bmatrix} = \mathbf{b},$$

where

$$\mathbf{A} = \begin{bmatrix} 2\mathbf{A}_{\mathrm{dif}}^{\mathsf{T}}\mathbf{A}_{\mathrm{dif}} & \begin{bmatrix} \mathbf{0} & \mathbf{A}_{\mathrm{dir}} \end{bmatrix}^{\mathsf{T}} \\ \begin{bmatrix} \mathbf{0} & \mathbf{A}_{\mathrm{dir}} \end{bmatrix} & \mathbf{0} \end{bmatrix} \in \mathbb{R}^{(b_\eta + 3b_s + 3) \times (b_\eta + 3b_s + 3)}$$

$$\mathbf{b} = \begin{bmatrix} \mathbf{0} \\ \mathbf{m}_{\mathrm{dir}} \end{bmatrix} \in \mathbb{R}^{b_\eta + 3b_s + 3}$$

The estimates of $\mathbf{x}^*$ and $\boldsymbol{\mu}^*$ is found by $\mathbf{A}^{-1}\mathbf{b}$ if $\mathbf{A}^{-1}$ exists. Since the system in Eq. (8) is already overdetermined and sufficiently robust, we found that the explicit positivity constraints are unnecessary, as confirmed by our experiments. The full estimation of pixel-to-wavelength mapping and camera spectral sensitivity takes about two to four minutes per camera on a single core of an Intel Xeon CPU.

G. Existing diffraction grating-based method

The previous method [1] uses diffracted observations from multiple light sources and images of a transmissive color chart. In theory, their method can be applied in combination with our estimation method of the pixel-to-wavelength mapping. Specifically, we can use a diffracted observation from a single light source and treat the observation of the direct light as a single patch of the transmissive color chart. However, based on our preliminary experiments, we found that their solution method based on stochastic gradient descent using the ADAM optimizer is unstable.

Figure S6 shows their estimation results by varying the number of patches in the transmissive color chart. As stated above, the proposed setup corresponds to the single-patch case, and we confirm that stable estimation with [1] requires using more than 10 patches.

As in the one-patch result from the above preliminary experiment, the estimation in the real-world experiments shows RE $= 0.58$ to 0.65 when evaluated on the same input as ours. This emphasizes the effectiveness of the proposed solution method using a linear solution method, which works well for the single-patch case.

H. Slit width impact

This section outlines the conditions under which one-to-one pixel-to-wavelength mapping can be assumed. One-to-one pixel-to-wavelength mapping is ensured by two conditions: a) diffraction resolvance $\Delta\lambda = \lambda_f / N$ (with N being the number of slits) determined by the Rayleigh criterion must be finer than the ground-truth resolvance $\Delta\lambda_{\mathrm{GT}}$, and b) the diffracted pattern must span at least f pixels, *i.e.*, $n \geq f$. Given $\Delta\lambda_{\mathrm{GT}} \approx 10\,\mathrm{nm}$ and $\lambda_f = 720\,\mathrm{nm}$ from CamSPECS XL, condition (a) holds when $N \geq 100$. Our gratings (500 or 1000 slits/mm) meet this, and we align the camera so that the pattern spans at least $f = 31$ pixels.

I. Consistency of efficiency estimation

To assess the consistency of the diffraction grating efficiency estimation across different scenes, we show a comparison of the estimated efficiencies. Figure S7 shows the estimated efficiencies for the scenes in the real-world ex-

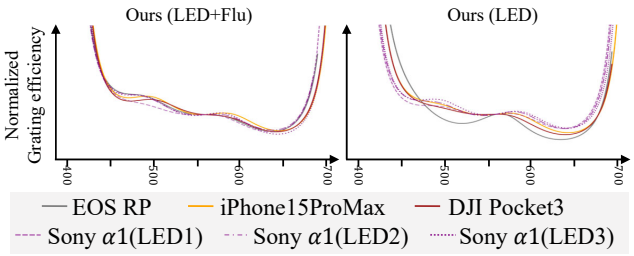

Figure S7. Normalized grating efficiency plot from Ours (LED+Flu) and Ours (LED).

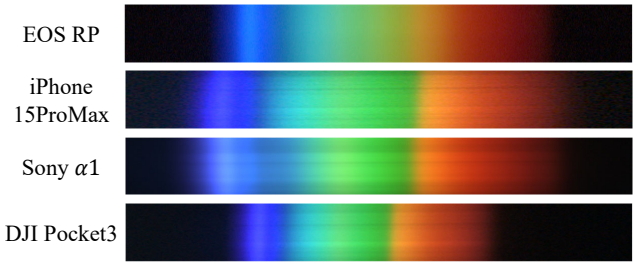

Figure S8. Diffraction observations in the real-world experiments.

periments (see Fig.7 in the main paper), which employ different cameras with varying relative positions between the camera and the slit, resulting in varying pixel-to-wavelength mappings. The mean cosine similarities of all estimations are 0.97 (LED+Flu) and 0.96 (LED), indicating the consistency of the proposed efficiency estimation method.

J. Diffracted light observation in the real-world experiments

Figure S8 shows the input diffracted light observations in the experiments. The observed diffraction patterns vary with each camera's spectral sensitivity, enabling us to estimate those sensitivities.

References

[1] Mikko E Toivonen and Arto Klami. Practical camera sensor spectral response and uncertainty estimation. *Journal of Imaging*, 6(8):79, 2020. 4