Compact Low-Cost Hyperspectral Imaging via Angular-to-Spectral Diversity Conversion — Supplementary Material —

Kazuma Fujiwara[†1], Takuya Funatomi[1,2], Kazuya Kitano[1], Yuki Fujimura[1], and Yasuhiro Mukaigawa[1]

[1] Nara Institute of Science and Technology, Nara, Japan
[2] Kyoto University, Kyoto, Japan
{fujiwara.kazuma.fj4, kitano.kazuya, fujimura.yuki, mukaigawa}@is.naist.jp, funatomi.takuya.2c@kyoto-u.ac.jp

In this supplementary material, we provide additional methodological and experimental details that support the main text. We first present the full preprocessing procedure, then provide an additional qualitative comparison with representative snapshot HSI systems, describe the detailed synthetic experimental setup, provide an experimental validation of the optical path difference (OPD) distribution using real data, and finally report the spatial-resolution evaluation using a USAF target.

1 Full Preprocessing Details

1.1 Measuring View-Channel-Wise Birefringent Filter Transmittance

We calibrate the view-channel-wise effective birefringent filter (BRF) transmittance $B_v(\lambda)$ using the setup shown in Figure 1. A monochromator (Optometrics, DMC1-03) equipped with a tungsten halogen light source generates narrowband illumination. Specifically, we use the tungsten halogen source (Optometrics, TS-110), which mounts directly to the monochromator. The monochromator output is split by a bifurcated fiber into two paths: one path is directed to our optical system, and the other is directed to a spectrometer. The spectrometer (Ocean Optics, Maya 2000) is used to verify the center wavelength of the generated light.

For this calibration, the objective lens was removed. Before entering the optical system, the monochromator output was passed through an external diffuser (Thorlabs, ED1-C20-MD) used only to spatially homogenize the illumination. The homogenized light then entered the system, passed through the system diffuser, the BRF, and the kaleidoscope, and was recorded by the camera.

We scan the monochromator at 10 nm intervals from 430 nm to 670 nm and capture one image at each wavelength. For each wavelength, we normalize each pixel in all replicated view-channels by the mean intensity of the BRF-free region in the central (zero-reflection) view-channel. The normalized responses over wavelength are treated as pixel-wise measurements of the effective transmittance

[†] Corresponding author.

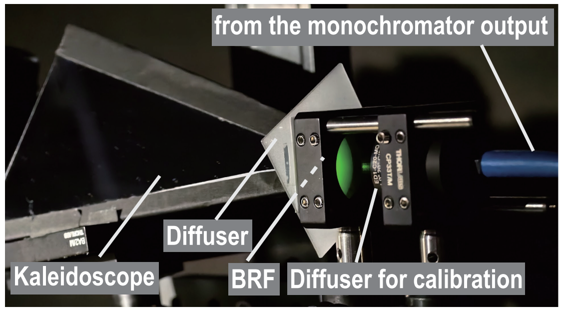


Fig. 1: Calibration setup for measuring the view-channel-wise BRF transmittance $B_v(\lambda)$.

profiles corresponding to $B_v(\lambda)$ and are used to construct the per-pixel sensing matrix.

1.2 Estimating Illumination–Sensor Response and Throughput Factors

After calibrating the view-channel-wise BRF transmittance $B_v(\lambda)$, we estimate the product of the illumination spectrum and the camera RGB sensitivities, $\phi^c(\lambda) = l(\lambda) f^c(\lambda)$, together with wavelength-independent throughput factors g_v, where v denotes the view-channel and $c \in \{R, G, B\}$ denotes the sensor color channel.

For a white reference target, the observation of view-channel v and color channel c is modeled as

$$I^c_{v,\mathrm{w}} = g_v \int_{\Lambda} \phi^c(\lambda)\, B_v(\lambda)\, d\lambda + \varepsilon^c_v, \tag{1}$$

where g_v is a wavelength-independent throughput factor, and ε^c_v denotes the residual term accounting for measurement noise and model mismatch. Discretizing Equation (1), we obtain

$$I^c_{v,\mathrm{w}} = g_v\, \mathbf{b}_v^\top \phi^c + \varepsilon^c_v, \tag{2}$$

where $\phi^c \in \mathbb{R}^{N_\lambda}$ is the discretized spectral response and $(\mathbf{b}_v)_m = B_v(\lambda_m)\Delta\lambda_m$.

With g_v fixed, we estimate ϕ^c by

$$(\hat{\phi}^c, \hat{\alpha}_c) = \arg\min_{\phi^c, \alpha_c} \sum_v \left(I^c_{v,\mathrm{w}} - g_v\, \mathbf{b}_v^\top \phi^c\right)^2 + \beta_\phi \|\mathbf{D}\phi^c\|_2^2 + \gamma_\mathrm{h} \|\phi^c - \alpha_c \tilde{f}^c\|_2^2, \tag{3}$$

where $\mathbf{D}$ is a difference matrix along wavelength, $\tilde{f}^c$ is an independently measured relative camera-sensitivity shape prior, and α_c absorbs the unknown scale of the prior. Here, the prior does not force $\phi^c(\lambda)$ to coincide with the bare camera sensitivity $f^c(\lambda)$. Rather, it encourages a similar spectral shape while allowing $\phi^c(\lambda)$ to absorb additional wavelength-dependent effects introduced by the imaging optics, including the diffuser and the objective lens. This is particularly necessary because the objective lens was removed during the calibration of $B_v(\lambda)$, so its wavelength-dependent response is not included in $B_v(\lambda)$ and is instead absorbed into $\phi^c(\lambda)$.

With ϕ^c fixed, we update the channel-wise throughput factors by regularized least squares and alternate these updates until convergence. In practice, we use multiple observations for each view-channel and allow mild spatial variation in the gain correction with additional smoothness regularization. Note that the absolute values of β_ϕ and γ_h are implementation-dependent. In our implementation, the observation term is evaluated on the measured intensity scale, and its numerical scale is further affected by the wavelength discretization and the scaling of the calibrated transmittance and sensitivity terms. Therefore, relatively large numerical values of β_ϕ and γ_h are required to balance the data term and the prior terms. We used $\beta_\phi = 8 \times 10^5$ and $\gamma_\mathrm{h} = 8 \times 10^5$, and ran the alternating optimization for 3 iterations.

The resulting $\hat{\phi}^c(\lambda)$ and g_v are then used to construct the sensing matrix for subsequent reconstruction.

1.3 Inter-Channel Registration

We register the replicated view-channels using images of an asymmetric circle-grid calibration target captured through the system. For each target image, we crop the triangular region corresponding to each view-channel using predefined triangle vertices obtained from a separate calibration image of the replicated views. For channels with mirror-induced parity reversals, we apply the corresponding flip before feature detection.

We use the zero-reflection view-channel as the reference view and denote it by $v = 1$. We then detect a 4×5 asymmetric circle grid in each cropped view using OpenCV's circle-grid detector with a blob detector. We aggregate the detected grid correspondences over multiple target images and estimate a homography from each view-channel to the reference view using random sample consensus (RANSAC). The reference view is used as the common coordinate system, and its homography is therefore set to the identity.

The estimated inter-channel homographies are fixed after calibration and used to warp all view-channels into the common reference coordinate system before forming the observation vector for spectral reconstruction.

2 Additional Comparison with Existing Snapshot HSI Systems

Table 1 summarizes how the proposed system differs from representative snapshot HSI systems in terms of form factor, optical implementation, add-on compatibility, and reconstruction requirements. The goal is not to exhaustively rank all systems, but to clarify the design trade-offs relevant to our target: a compact, low-cost, off-the-shelf camera add-on usable without supervised training.

Table 1: Qualitative comparison with representative snapshot HSI systems. "Size / form factor" lists reported dimensions when available, otherwise the reported prototype form factor; the Fujimoto et al. length is author-measured. "Off-the-shelf," "Camera add-on," and "Train-free" denote spectral-coding optics used as off-the-shelf components without custom fabrication, single-camera attachment without multi-camera or sensor-level modification, and reconstruction without supervised training, respectively.

Method	Size / form factor	Optical strategy	Off-the-shelf	Camera add-on	Train-free	Trade-off / note
Baek et al. (2017) [1]	DSLR + small front prism; dimensions not reported	Prism dispersion over edges	Yes	Yes	Yes	Very compact front attachment, but relies on sparse edge cues and costly optimization.
Oh et al. (2016) [8]	Three consumer cameras	RGB cameras with distinct spectral sensitivities	Yes	No	Yes	Low-cost cameras, but requires multi-camera registration; dynamic scenes would need synchronized capture.
Takatani et al. (2017) [9]	Filter-attached faced mirrors: 300 mm length, 10 mm spacing	Color-filter-attached faced reflectors	Yes	Yes	Yes	Very low-cost mirror attachment, but assumes flat scenes and repeated reflections reduce spatial resolution and SNR.
Manakov et al. (2013) [7]	Prototype: ∼1000 mm incl. camera/main lens; multiplier: $36 \times 24 \times 300$ mm	Kaleidoscope + filters + relay optics	Partly	Yes	Yes	Reconfigurable single-camera add-on, but relay optics increase length.
Zhao et al. (2019) [11]	Relay prototype; compact sensor-mask version proposed	Random printed mask + relay optics	No	Partly	No	Low-cost printed mask, but requires custom mask printing and CNN-based reconstruction.
Fujimoto et al. (2025) [3]	Relay prototype: ∼50 cm total length	Natural birefringent thin-section filter	No	Partly	Yes	No microfabrication, but reconstruction assumes locally uniform patches, reducing spatial resolution.
Gorman et al. (2010) [4]	Cage-mounted birefringent demultiplexer; system length not reported	Cascaded birefringent interferometers + Wollaston demultiplexing	No	Partly	Yes	Direct spectral-image recording, but wider FoV/NA requires large, costly Wollaston-prism optics.
Kudenov et al. (2012) [6]	$15 \times 15 \times 10\,\mathrm{mm}^3$ common-path interferometer; benchtop prototype	Nomarski-prism interferometer + lenslet arrays	No	No	Yes	Compact core optics, but specialized interferometric optics and multi-step calibration are required.
DOE/filter-array systems [2, 5, 10]	On-chip / sensor-level coding elements	Custom DOE or filter-array coding	No	No	Varies	Compact, but relies on fabricated coding elements or sensor-level modification.
Ours	**21 cm add-on; 40 cm incl. camera**	**Tapered kaleidoscope + directly attached BRF**	**Yes**	**Yes**	**Yes**	**No relay/cascade; directly attached off-the-shelf optics, but reduced reconstruction area from view replication.**

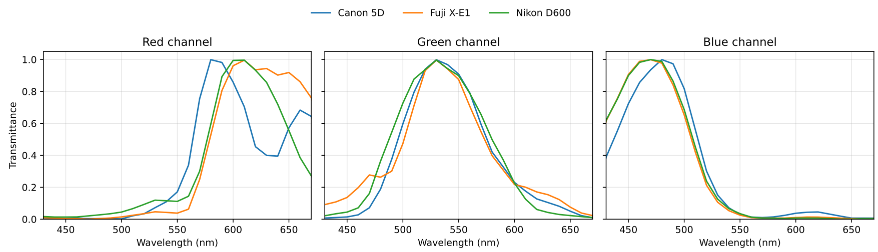


Fig. 2: Camera spectral sensitivities used to model the *3Camera* baseline [8].

3 Detailed Synthetic Experimental Setup

We compare our method with two low-cost snapshot baselines, *3Camera* [8] and *Faced Reflectors* [9]. For all methods, the method-specific channel-wise spectral responses are used only in the forward model to generate synthetic observations, while spectral reconstruction is performed using the same non-negative smoothness-regularized least-squares formulation. This protocol is intended to isolate the effect of the spectral responses used for measurement formation.

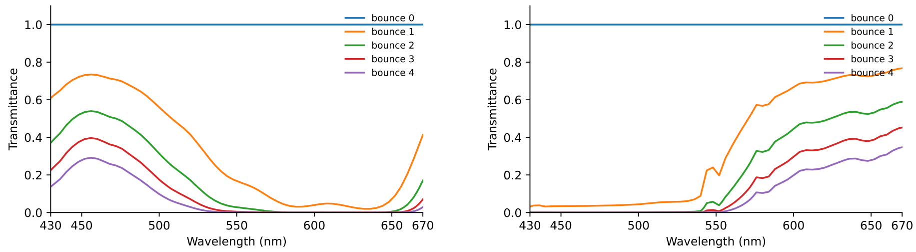
We generate synthetic observations from the 24 ColorChecker reflectance spectra on the same wavelength grid used in our system, namely 430 nm to 670 nm at 10 nm intervals (25 wavelength samples). To compare the methods under matched signal levels, we apply exposure matching so that, for each method, the maximum noise-free observation over all 24 patches and all observation entries is approximately 255 digital numbers (DN). After this exposure matching, we add i.i.d. Gaussian noise in the observation domain.

For each tested noise level, we sweep the smoothness weight β over 0 and logarithmically spaced values from 10^{-2} to 10^{6}, and report the best peak signal-to-noise ratio (PSNR) over β. This evaluation is intended to compare the sensing designs under a common reconstruction model rather than to tune reconstruction settings separately for each method. As in Section 1.2, the absolute value of the smoothness weight is implementation-dependent. In our implementation, the observation vector is represented on an exposure-matched DN scale, and its numerical scale is further affected by the wavelength discretization and by the scaling of the spectral response terms in the sensing matrix. Therefore, relatively large numerical values of β are required to balance the regularization term.

We next describe the spectral responses used for each method in the forward model. For the baseline methods, the spectral responses used in the forward model are digitized from spectral plots reported in the original papers.

3.1 3Camera Baseline

For the *3Camera* baseline, we use the three-camera setting reported in [8]: Canon 5D, Fuji X-E1, and Nikon D600. Each camera-color pair is treated as an indepen-



(a) #67 Light Sky Blue. **(b)** Thin clear orange filter.

Fig. 3: Bounce-dependent effective transmittances used to model the *Faced Reflectors* baseline [9].

dent measurement channel, yielding nine channels in total. The corresponding spectral sensitivities are shown in Figure 2.

Under the common linear model, the effective spectral response of camera q and color channel c is

$$T_{q,c}(\lambda) = f_q^c(\lambda), \tag{4}$$

where $q \in \{1, 2, 3\}$ indexes the cameras and $c \in \{R, G, B\}$ indexes the sensor color channels.

3.2 Faced Reflectors Baseline

For the *Faced Reflectors* baseline, we follow the original bounce-dependent model in which the same filter is attached to the opposed reflectors. If the base filter transmittance is $\tau(\lambda)$, the i-th bounce is modeled by

$$T_{i,c}(\lambda) = f^c(\lambda)\, \tau(\lambda)^i, \tag{5}$$

where $f^c(\lambda)$ is the RGB sensor sensitivity and i is the bounce order.

In the synthetic evaluation, we use two filter settings from [9]: #67 Light Sky Blue, which was reported to provide favorable conditioning in the synthetic analysis, and the thin clear orange filter, which was used in their real-data prototype. Figure 3 shows the corresponding bounce-dependent transmittances used in the forward model.

4 Experimental Validation of the OPD Distribution Using Real Data

Figure 4 compares the measured OPD estimates with the simulated view-channel geometry and shows representative transmittance spectra. In the main text, the angular distribution of the OPD is analyzed primarily through simulation. Here, we provide an experimental validation showing that the OPD distribution observed in the real system is qualitatively consistent with the simulated prediction.

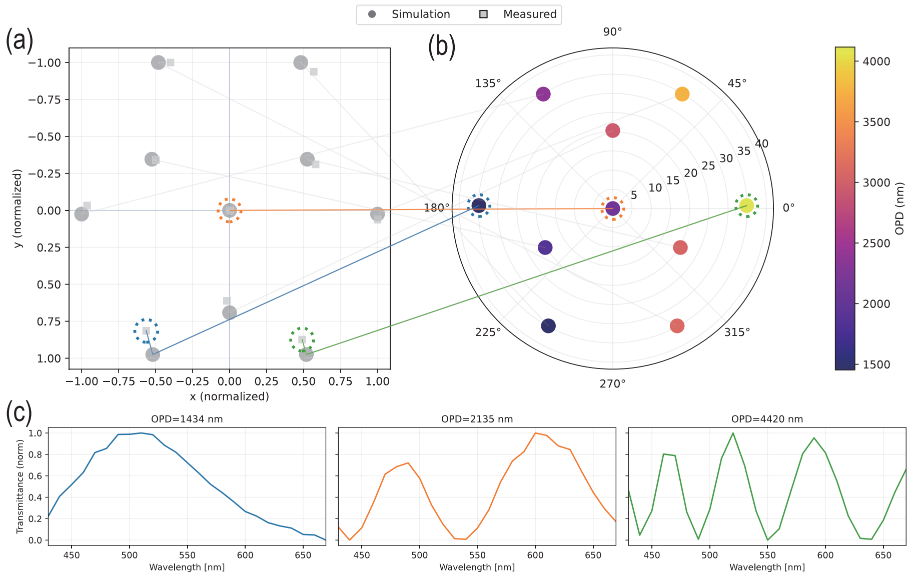


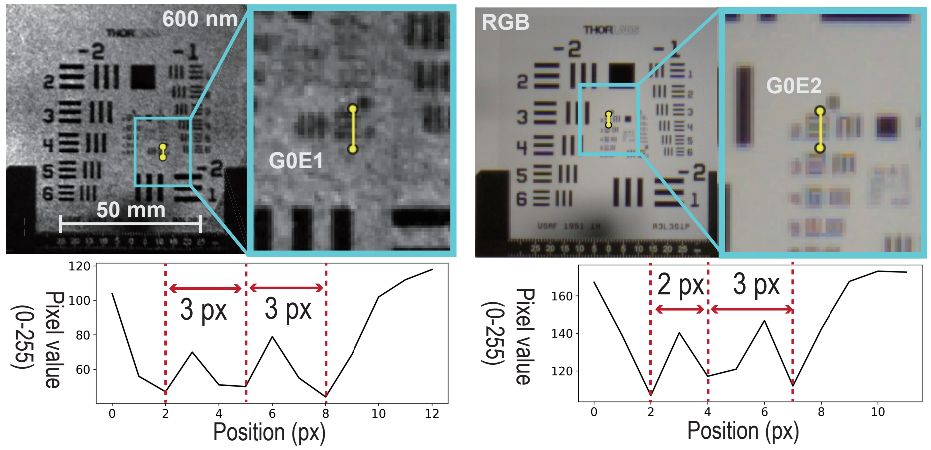
Fig. 4: Validation of the OPD, Δ_{o-e}, distribution from measured data. (a) Correspondence between measured points (squares) and simulated view-channels (circles) on the image plane. Each matched pair is connected by a line segment. (b) Estimated OPD values visualized on the (α, σ) polar parameterization of the view-channels, where color indicates the estimated Δ_{o-e}. Matched channel pairs are again connected by line segments. (c) Examples of normalized transmittance spectra at the three dashed-circle locations highlighted in (a) and (b), corresponding to small, intermediate, and large OPD values.

For each measured view-channel, we use the calibrated spectral transmittance profile and estimate the OPD from its periodic modulation. Because the birefringence-induced transmittance is periodic with respect to wavenumber rather than wavelength, we first convert the spectrum from wavelength λ to wavenumber $k = 1/\lambda$ and resample it on a uniformly spaced grid in k. Under the transmittance model in the main text, the view-channel-wise spectral modulation can be written as

$$T(k) \propto \frac{1}{2} \left[1 - \cos(2\pi \, \Delta_{o-e} \, k)\right], \qquad k = \frac{1}{\lambda}, \tag{6}$$

so that the dominant frequency in the wavenumber domain directly corresponds to the OPD Δ_{o-e}.

We therefore estimate Δ_{o-e} for each measured point by applying a 1D fast Fourier transform (FFT) to the uniformly resampled transmittance profile in the wavenumber domain and identifying the dominant spectral frequency component. Independently, for each simulated view-channel, we compute its image-plane location and angular coordinates (α, σ) by ray tracing under the same



(a) Reconstructed HSI at 600 nm. **(b)** Without kaleidoscope/diffuser.

Fig. 5: Spatial-resolution evaluation using a 1951 USAF target. (a) Reconstructed HSI result at 600 nm. (b) RGB reference captured without the kaleidoscope/diffuser under the same objective lens, camera, and image-plane scale.

tapered-kaleidoscope geometry as in the main text. To compare simulation and measurement, we match each measured point to the nearest simulated view-channel on the image plane and transfer the estimated OPD values to the corresponding simulated angular coordinates (α, σ).

The measured OPD distribution is qualitatively consistent with the simulated prediction. In particular, it exhibits approximate mirror symmetry and varies across view-channels. These observations indicate that the replicated measurements provide angle-dependent spectral diversity in the real system. Although this validation is qualitative and based on discrete channel samples, it remains consistent with the optical design assumption underlying the proposed system.

5 Spatial Resolution Evaluation

This section provides additional details for the spatial-resolution evaluation summarized in the main paper. We evaluate the effective spatial resolution of the prototype using a 1951 USAF resolution target (Thorlabs R3L3S1P). Figure 5 compares a representative reconstructed HSI band at 600 nm with an RGB reference captured without the kaleidoscope/diffuser.

For the proposed-system measurement, the target was captured with the same prototype, calibration, inter-channel registration, reconstruction pipeline, and smoothness weight used for the real-scene reconstruction in the main paper. The capture settings were 0.25 s, ISO 200, and F/8 under daylight-level artificial solar illumination. We reconstructed 25 bands over 430 nm to 670 nm at 10 nm intervals and evaluated the highest resolvable USAF group/element patterns.

For comparison, the RGB reference image was captured without the kaleidoscope/diffuser using the same objective lens and camera. The object distance and image-plane scale were matched between the reconstructed HSI result and the RGB reference. The RGB reference was captured at 0.01 s, ISO 200, and F/8.

The 1951 USAF target defines the spatial frequency of group G and element E as

$$r(G, E) = 2^{G+(E-1)/6} \quad [\text{lp/mm}]. \tag{7}$$

Most reconstructed HSI bands show resolvable structures at 0.89 lp/mm to 1.00 lp/mm, corresponding to Group −1 Element 6 to Group 0 Element 1. The representative 600 nm band in Figure 5a is consistent with this range. The RGB reference without the kaleidoscope/diffuser in Figure 5b resolves 1.12 lp/mm, corresponding to Group 0 Element 2. Thus, in this USAF test, the limiting resolution of the reconstructed HSI is within one to two USAF group/element steps of the RGB reference. This indicates that the dominant spatial cost of the current prototype is the reduced valid reconstruction area from 10-view replication, rather than a large additional limiting-resolution loss caused by diffuser blur or inter-channel registration.

Valid reconstruction region. We define the valid reconstruction region as pixels with at least three registered BRF-modulated view-channel observations. Near the boundaries of the registered view-channel regions, not all ten view channels are available at every pixel. For such pixels, unavailable observations are masked out, and the same pixel-wise reconstruction problem is solved using the corresponding available rows of the sensing matrix. In the 10-view configuration, this valid region contains approximately 3.24×10^5 pixels, corresponding to 4.5 % of the bounding box enclosing the registered view-channel regions.

References

1. Baek, S.H., Kim, I., Gutierrez, D., Kim, M.H.: Compact single-shot hyperspectral imaging using a prism. ACM Trans. Graph. **36**(6) (Nov 2017). `https://doi.org/10.1145/3130800.3130896`
2. Bian, L., Wang, Z., Zhang, Y., Li, L., Zhang, Y., Yang, C., Fang, W., Zhao, J., Zhu, C., Meng, Q., Peng, X., Zhang, J.: A broadband hyperspectral image sensor with high spatio-temporal resolution. Nature **635**(8037), 73–81 (Nov 2024). `https://doi.org/10.1038/s41586-024-08109-1`
3. Fujimoto, Y., Kitano, K., Fujiwara, K., Fujimura, Y., Funatomi, T., Mukaigawa, Y.: Snapshot hyperspectral imaging using petrographic thin section. In: International Conference on Machine Vision Applications (MVA) 2025 (Jul 2025). `https://doi.org/10.34385/proc.93.P2-4`, session P2-4, Publication date: 2025-07-26
4. Gorman, A., Fletcher-Holmes, D.W., Harvey, A.R.: Generalization of the lyot filter and its application to snapshot spectral imaging. Opt. Express **18**(6), 5602–5608 (Mar 2010). `https://doi.org/10.1364/OE.18.005602`
5. Jeon, D.S., Baek, S.H., Yi, S., Fu, Q., Dun, X., Heidrich, W., Kim, M.H.: Compact snapshot hyperspectral imaging with diffracted rotation. ACM Transactions on

Graphics (Proc. SIGGRAPH 2019) **38**(4), 117:1–13 (2019). https://doi.org/10.1145/3306346.3322946
6. Kudenov, M.W., Dereniak, E.L.: Compact real-time birefringent imaging spectrometer. Opt. Express **20**(16), 17973–17986 (Jul 2012). https://doi.org/10.1364/OE.20.017973
7. Manakov, A., Restrepo, J.F., Klehm, O., Hegedüs, R., Eisemann, E., Seidel, H.P., Ihrke, I.: A reconfigurable camera add-on for high dynamic range, multispectral, polarization, and light-field imaging. ACM Trans. Graph. **32**(4) (Jul 2013). https://doi.org/10.1145/2461912.2461937
8. Oh, S.W., Brown, M.S., Pollefeys, M., Kim, S.J.: Do it yourself hyperspectral imaging with everyday digital cameras. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2461–2469 (2016). https://doi.org/10.1109/CVPR.2016.270
9. Takatani, T., Aoto, T., Mukaigawa, Y.: One-shot hyperspectral imaging using faced reflectors. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4039–4047 (2017)
10. Yako, M., Yamaoka, Y., Kiyohara, T., Hosokawa, C., Noda, A., Tack, K., Spooren, N., Hirasawa, T., Ishikawa, A.: Video-rate hyperspectral camera based on a cmos-compatible random array of fabry–pérot filters. Nature Photonics **17**(3), 218–223 (Mar 2023). https://doi.org/10.1038/s41566-022-01141-5
11. Zhao, Y., Guo, H., Ma, Z., Cao, X., Yue, T., Hu, X.: Hyperspectral imaging with random printed mask. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10149–10157 (2019)