Letter

Video-rate hyperspectral camera based on a CMOS-compatible random array of Fabry–Pérot filters

Motoki Yako [1], Yoshikazu Yamaoka[1], Takayuki Kiyohara [1], Chikai Hosokawa[1], Akihiro Noda [1], Klaas Tack [2], Nick Spooren[2], Taku Hirasawa [1] & Atsushi Ishikawa [1]✉

Hyperspectral (HS) imaging provides rich spatial and spectral information and extends image inspection beyond human perception. Existing approaches, however, suffer from several drawbacks such as low sensitivity, resolution and/or frame rate, which confines HS cameras to scientific laboratories. Here we develop a video-rate HS camera capable of collecting spectral information on real-world scenes with sensitivities and spatial resolutions comparable with those of a typical RGB camera. Our camera uses compressive sensing, whereby spatial–spectral encoding is achieved with an array of 64 complementary metal–oxide–semiconductor (CMOS)-compatible Fabry–Pérot filters placed onto a monochromatic image sensor. The array affords high optical transmission while minimizing the reconstruction error in subsequent iterative image reconstruction. The experimentally measured sensitivity of 45% for visible light, the spatial resolution of 3 px for 3 dB contrast, and the frame rate of 32.3 fps at VGA resolution meet the requirements for practical use. For further acceleration, we show that AI-based image reconstruction affords operation at 34.4 fps and full high-definition resolution. By enabling practical sensitivity, resolution and frame rate together with compact size and data compression, our HS camera holds great promise for the adoption of HS technology in real-world scenarios, including consumer applications such as smartphones and drones.

Hyperspectral (HS) cameras—owing to their ability to provide rich spectral information at each pixel of an image—have stimulated advanced sensing applications, including machine vision, food and agricultural analysis, environmental monitoring and healthcare[1–13]. Although three general types of HS cameras (spatial-scanning[14], spectral-scanning[15] and snapshot[16,17]) are commercially available, their adoption in industrial and consumer applications is hindered by fundamental drawbacks such as low sensitivity, resolution and/or frame rate compared with an RGB camera (Supplementary Section 1). The emerging field of computational imaging has the potential to enhance the sensing capability of

HS cameras by optimal hardware design and image post-processing. Compressed sensing is a technique to efficiently acquire signals from under-sampled measurements, and has contributed to improving performance in many sensing systems[18]. Recent demonstrations include an on-chip spectroscopic device[19–26] and a coded aperture snapshot spectral imager[27–31]. However, no video-rate HS camera has been made with RGB camera-comparable sensitivity and resolution, thus hindering the potential of HS in industrial and consumer applications.

In this Article, we demonstrate a HS camera that exploits efficient acquisition of the spatial and spectral information compressed in

[1]Panasonic Holdings Corporation, Kadoma, Japan. [2]Imec, Leuven, Belgium. ✉e-mail: ishikawa.atsushi@jp.panasonic.com
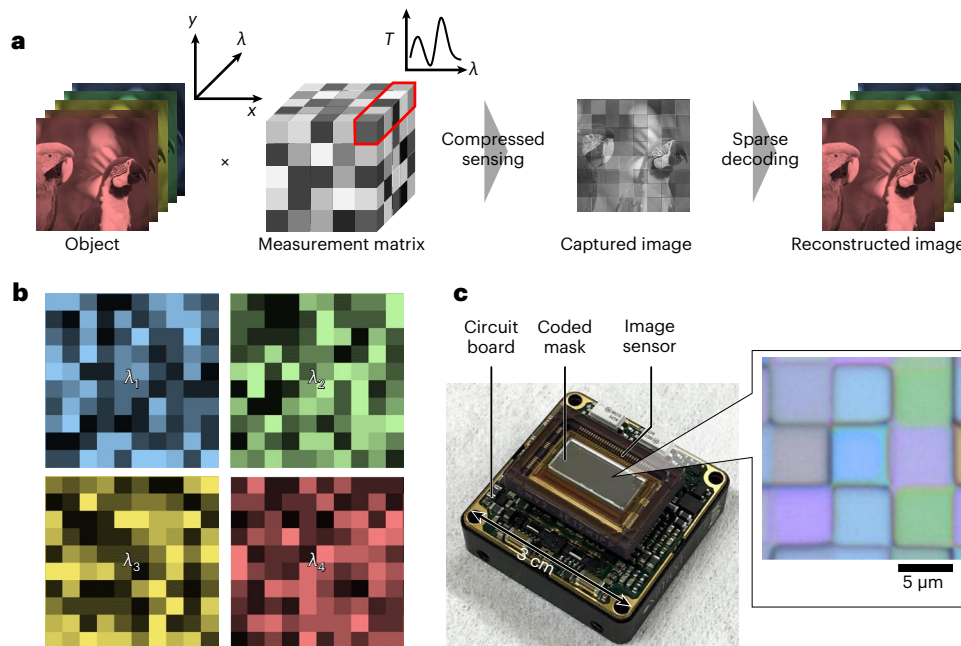
Fig. 1 | Hyperspectral camera using a spatial–spectral coded mask. a, Conceptual scheme of the compressive HS sensing. Light from an object is spatially and spectrally encoded by a measurement matrix and captured as a monochromatic image, followed by image reconstruction based on spatial sparsity of the object. **b,** Conceptual images of transmittance patterns of a coded mask at four different wavelengths to realize the measurement matrix.

The coded mask shows spatially random transmittance patterns with a small correlation between different wavelength bands, thereby realizing spatially and spectrally random sampling. **c,** A photograph of our HS camera with the coded mask directly placed onto the surface of a monochromatic image sensor. The inset shows a microscope image of the mask, which is an array of square cells with a pixel pitch of 5.5 μm, matched with that of the image sensor.

two-dimensional sensor signals. Our camera shows a sensitivity of 45% for visible light, a spatial resolution of 3 px for 3 dB contrast, and a frame rate of 32.3 fps at VGA resolution (0.3 MP). These metrics are comparable with an equivalent RGB camera and thus meet the requirements for practical use.

In our HS imaging method (Fig. 1a), light from an object is spatially and spectrally encoded by a measurement matrix, followed by image reconstruction based on spatial sparsity of the object. The measurement matrix is implemented as a coded mask of spatially random transmittance patterns with a small correlation between different wavelengths, that is, spectral randomness (conceptually shown in Fig. 1b and Supplementary Fig. 1). As shown in Fig. 1c, the coded mask is directly placed onto a monochromatic image sensor, otherwise, for mass production, it can be monolithically integrated onto the sensor using a complementary metal–oxide–semiconductor (CMOS)-compatible process (see Methods). The mask is an array of square cells with a pixel pitch of 5.5 μm (inset of Fig. 1c), matched with that of the image sensor.

Encoded light from the object is thus captured as a single monochromatic image (compressed image), and image reconstruction is performed assuming the spatial sparsity of the object. The compressed sensing process and the image reconstruction under the sparsity assumption can be expressed as follows:

$$g(x,\ y) = \sum_{i=1}^{N} A_i(x,y) f_i(x,y) \qquad (1)$$

$$Q(f) = \arg\min_{f} \left\{ \left\| g - \sum_{i=1}^{N} A_i f_i \right\|_2^2 + \tau \left\| \mathrm{TV}(f) \right\|_1 \right\} \qquad (2)$$

where N is the number of wavelength bands, g is the compressed image, A is the measurement matrix based on the transmittance patterns of

the coded mask, f is the reconstructed image, τ is the regularization coefficient and TV represents total variation. In this work, equation (2) is solved by an iterative method based on a two-step iterative shrinkage thresholding (TwIST) algorithm[32]. Solving equation (2) minimizes the cost function Q(f) to find the reconstructed image f through a TV denoising (regularization) process, with τ compared with the first L$_2$-norm fidelity term. A smaller value of τ performs weaker TV denoising to maintain the sharp edges of the reconstructed image, but makes the reconstruction process more unstable against noise on the compressed image; the spatial resolution is thus essentially limited by sensor noise[33].

Although equation (2) is an underdetermined linear system, its convergence is mathematically ensured with a non-zero reconstruction error if A is incoherent for the object signal to be measured (Supplementary Section 2). In this case, incoherence of the measurement matrix is achieved by a spatially and spectrally random transmittance of a coded mask. As shown in Fig. 1b, the mask is designed to show spatially random transmittance patterns with a small correlation between different wavelength bands, that is, spectral randomness.

To design an incoherent coded mask for low-error image reconstruction, we started by conducting numerical simulations to study the reconstruction error as a function of the spatial and spectral randomness of the mask (Supplementary Section 2). Virtual coded masks were numerically generated with varying spatial (spectral) randomness while ensuring the perfect randomness in the spectral (spatial) dimension so that the spatial (spectral) randomness was separately evaluated. The reconstruction error arises not only from the mask's lack of spatial and spectral randomness, but also from noise on the compressed image. Our goal here was therefore to design the mask that produces a maximum reconstruction error of 2%, given that the image sensor used in this work (ams, CMV2000-3E12M1PP) shows a signal fluctuation around 2% (Supplementary Fig. 8).

The spatial randomness of the mask was characterized for each wavelength band using an evaluation index σ/μ, where the standard
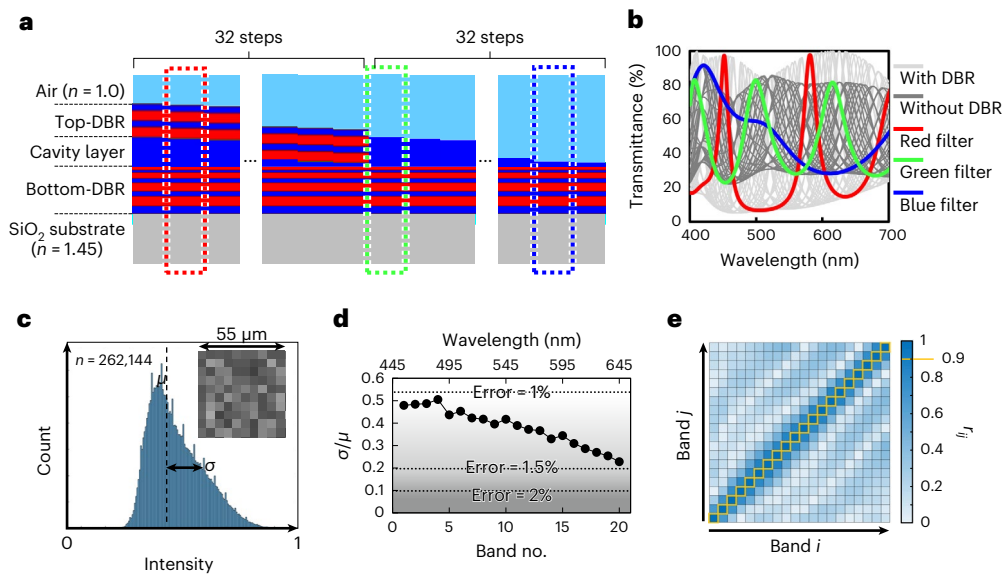
Fig. 2 | Structures and characteristics of Fabry–Pérot filters in the coded mask. a, Schematic cross-section of the Fabry–Pérot filters, square cells in the coded mask consisting of DBRs and a cavity layer. The cavity thickness is progressively varied to produce unique transmission spectra. Half of the filters (32) have no top DBR to improve spectral randomness over the visible wavelength range. **b**, Typical transmission spectra of the 64 Fabry–Pérot filters. The red, green and blue spectra correspond to the filters indicated in **a**. **c**, A transmittance histogram of the coded mask (512 × 512 = 262,144 cells) at λ = 550 nm and the image of a 10 × 10 px area (inset). The horizontal axis indicates signal intensity divided by the upper limit (8 bit = 255 in this work), whereas the vertical axis indicates pixel counts; σ and μ are derived from the histogram. **d**, Values of σ/μ (an evaluation index of the spatial randomness) for each wavelength band. The simulated reconstruction errors associated with the spatial randomness are indicated by the dotted lines. **e**, Two-dimensional correlation coefficients (r_{ij}) map of the coded mask between the wavelength bands i and j (where i and j are from 1 to 20, corresponding to λ = 450–650 nm in 10 nm steps). The smaller non-diagonal components ensure a smaller reconstruction error, whereas the diagonal components are equal to 1.

deviation (σ) and the average transmittance (μ) are obtained from the histogram of the transmittance pattern. As shown in Supplementary Fig. 2, the reconstruction error monotonically decreases as σ/μ increases and becomes smaller than 2% for $\sigma/\mu > 0.1$.

The spectral randomness of the mask was instead characterized by calculating the two-dimensional correlation coefficients between the wavelength bands i and j as:

$$r_{ij} = \left| \frac{\sum_m \sum_n (i_{mn} - \bar{i})(j_{mn} - \bar{j})}{\sqrt{\left(\sum_m \sum_n (i_{mn} - \bar{i})^2\right)\left(\sum_m \sum_n (j_{mn} - \bar{j})^2\right)}} \right| \quad (3)$$

where $i(j)_{mn}$ is the cell transmittance at location $(x, y) = (m, n)$ in the mask and $\bar{i}(\bar{j})$ is the average transmittance. The non-diagonal components ($i \neq j$) represent indices for the spectral randomness whereas the diagonal components ($i = j$) are naturally equal to 1. As shown in Supplementary Fig. 3, the reconstruction error is suppressed at ~1% for $r_{ij} < 0.9$ ($i \neq j$).

To realize a coded mask that satisfies $\sigma/\mu > 0.1$ and $r_{ij} < 0.9$ ($i \neq j$), and ensure a reconstruction error below 2%, Fabry–Pérot filters are used to form the square cells in the mask. Figure 2a shows the schematic cross-section of the Fabry–Pérot filters, which consist of distributed Bragg reflectors (DBRs) and a cavity layer on a SiO₂ substrate. Considering the required randomness and fabrication process complexity, 64 Fabry–Pérot filters are designed; half (32) of the filters have top and bottom DBRs with a cavity layer in between, whereas the other half (32) have only a bottom DBR with a cavity layer. By progressively varying the cavity thickness in 32 steps, the 64 filters show unique transmission spectra. More filters (for example, 128) would only result in a minor improvement to the spectral resolution, but such a design dramatically increases the difficulty of the fabrication process (Supplementary Fig. 5).

In Fig. 2b, the light (dark) grey curves show the simulated spectra of the Fabry–Pérot filters with (without) top DBRs, whereas the red, green and blue spectra correspond to the filters highlighted in Fig. 2a. The average transmittance of the 64 filters is 46.4% for the visible wavelength range (from λ = 450 nm to 650 nm). The coded mask comprising the 64 filters was then fabricated and integrated onto the image sensor, and the transmittance patterns were measured to determine the measurement matrix through a calibration process (see Methods). In the calibration, the wavelength bands were set from λ = 450 nm to 650 nm with a sampling interval $\Delta\lambda$ of 10 nm (a total of 20 bands).

Figure 2c shows an experimentally measured histogram and transmittance pattern (inset) of the fabricated mask at λ = 550 nm. The σ/μ values are plotted in Fig. 2d, where σ/μ gradually decreases at longer wavelengths because the Fabry–Pérot filters show broader transmission peaks leading to larger μ and smaller σ. The measured σ/μ is larger than 0.2 for all bands, corresponding to a reconstruction error of 1–1.5% (Supplementary Fig. 2).

Figure 2e shows the experimentally measured r_{ij} of the fabricated mask, where the yellow line indicates a contour at r_{ij} = 0.9. The fabricated mask satisfies $r_{ij} < 0.9$ ($i \neq j$), corresponding to ~1% reconstruction error, as expected from Supplementary Fig. 3.

We next evaluate the sensitivity of our HS camera by imaging objects under average office lighting with 550 lux, which is a relatively dim environment compared with natural scenes. Figure 3a,b shows photographs captured by our HS camera and a standard RGB camera (see Methods). The HS image obtained by our HS camera is rendered as a synthetic RGB image, showing a different hue from that of the RGB camera. The three expanded images in Fig. 3a,b indicate that the image quality of our HS camera is qualitatively comparable with that of the RGB camera. Although our HS camera loses spatial information owing to TV denoising, the reconstruction error at the edges is not clearly observed. A more detailed investigation on object boundaries
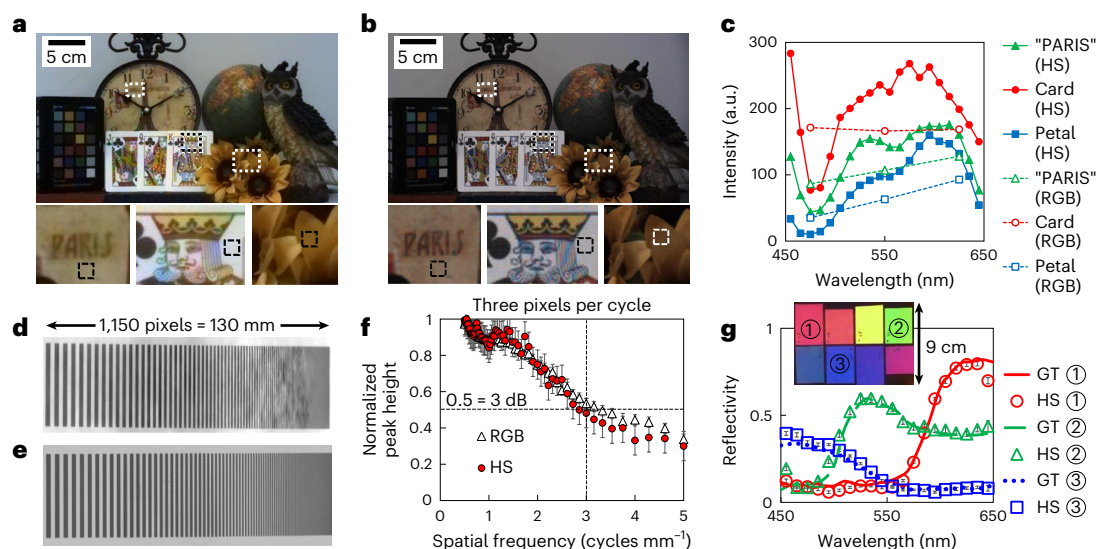
Fig. 3 | Experimental results of HS image acquisition. a,b, Photographs captured by our HS camera (**a**) and a standard RGB camera (**b**) under average office lighting with 550 lux. The HS image is rendered as a synthetic RGB image for clarity, showing a different hue from that of the RGB camera. Three expanded images indicate that the image quality of our HS camera is qualitatively comparable with the RGB camera. The average filter transmittances calculated from raw images are 45% for our HS camera and 37% for the RGB camera over $\lambda = 450$–650 nm (Supplementary Section 5). **c**, Signal intensity in each band obtained from small areas (10×10 px; shown as doted squares) in the expanded images of our HS camera and the RGB camera. Our HS camera shows an intensity dip of LED lighting at $\lambda = 485$ nm by obtaining spectral information of 20 bands, whereas the RGB camera does not show the spectral dip as it obtains only three bands. **d**, A ladder resolution chart taken by our HS camera (**d**) and the RGB camera (**e**). **f**, MTF curves calculated from the resolution chart. Both cameras show 3 dB contrast around the spatial frequency of three cycles per millimetre, corresponding to a line width of 3 px. Our HS camera shows hazy patterns in the area in which the line width is too narrow, probably due to the image reconstruction assuming spatial sparsity (Supplementary Section 7). The error bars indicate the s.d., which are obtained by calculating 150 MTF curves by changing the vertical positions in **d** and **e**. **g**, Spectra of standard colour samples (50×50 px for each colour sample) measured by our HS camera (points) and a spectral-scanning HS camera (ground truth, GT; shown as solid/doted lines). The error bars indicate s.d. of the signal intensity obtained from the pixels in each small square, showing that spatial signal fluctuation is negligible. A good spectral agreement was obtained with an absolute average error of 2.2% over 20 bands (Supplementary Section 9). White balance was performed in **g** to compensate the spectrum of LED lighting.

is performed below (Fig. 4 and Supplementary Section 7). Quantitatively, the average transmittance (that is, the effective sensitivity) is evaluated with reference to the monochromatic camera transmittance of 100%, as the monochromatic camera uses no spectral filter. The effective sensitivities are 45% for our HS camera and 37% for the RGB camera over the 450–650 nm wavelength range, much higher than that of a conventional HS camera (typically ~5%); the raw images used to calculate the sensitivities are shown in Supplementary Fig. 11. The experimental sensitivity of our HS camera is close to the simulation result (46.4%), comparable with that of the RGB camera. From three expanded images in Fig. 3a,b, signal intensity is obtained in each band and plotted in Fig. 3c. Our HS camera shows spectral information of 20 bands, whereas the RGB camera shows only three bands. An intensity dip of LED lighting at $\lambda = 485$ nm is clearly observed by our HS camera, which is not detected by the RGB camera. As demonstrated below (Fig. 4 and Supplementary Video 1), our HS camera enables video-rate (> 30 fps) acquisition of 20 bands of spectral information.

The spatial resolution is experimentally visualized in Fig. 3d,e and evaluated as modulation transfer function (MTF) curves in Fig. 3f by imaging a ladder resolution chart. The MTF curves indicate that the spatial resolution of our HS camera is comparable with the RGB camera, exhibiting 3 dB contrast (normalized peak height of 0.5) obtained at the spatial frequency of three cycles per millimetre. The spatial resolution of three cycles per millimetre indicates that 3 px are required to separate adjacent lines with 3 dB contrast, as the millimetre/pixel conversion ratio is 1,150 px/130 mm, as evidenced from Fig. 3d,e. This result is reasonable because edge smoothing occurs in our HS camera during the TV denoising process, whereas the RGB camera with a Bayer filter mosaic also loses spatial information due to 2×2 px convolution and inter-pixel crosstalk. Further evaluation of the spatial resolution using a star resolution chart is performed in Supplementary Section 8.

The spectral accuracy of our HS camera is experimentally evaluated by using standard colour samples, as shown in Fig. 3g. The continuous lines in Fig. 3g represent ground truth spectra obtained using a spectral-scanning HS camera (see Methods), whereas the symbols represent three typical spectra obtained by our HS camera. Note that white balance was performed for the spectra in Fig. 3g by using a diffuse reflectance target (SphereOptics, SG3151) to compensate the spectrum of LED lighting. The results are in good agreement, as evidenced by further testing all eight colours and obtaining an absolute average error of 2.2% over 20 bands (Supplementary Section 9). The experimentally measured error is a reasonable value considering the sensor noise (2%) and the simulated reconstruction error arising from the spatial and spectral randomness (1–1.5% and ~1%).

To further characterize the performance of the image reconstruction, numerical simulations were performed using HS image datasets (CAVE[34], Tokyo Tech[35] and Manchester[36]), as shown in Fig. 4. In the simulation, the compressed image was first synthesized from each dataset following equation (1) with the experimentally obtained measurement matrix; the corresponding HS images were then obtained. Here, no signal fluctuation was assumed because the signal fluctuation arises from an image sensor, but not from the compression/reconstruction processes. The impact of sensor noise on the spectroscopic performance and image quality is separately discussed in Supplementary Section 5.

We first evaluated the image reconstruction quality. The rows of Fig. 4a show the ground truth images, the reconstructed images and the absolute error maps simulated using the datasets (CAVE, Tokyo Tech and Manchester, along the columns). The ground truth and
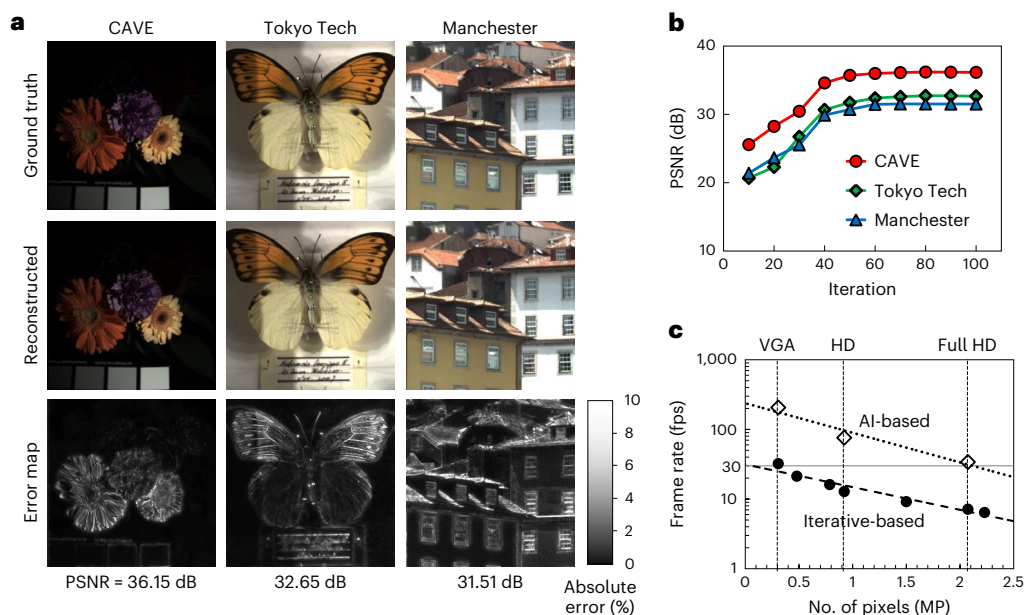
Fig. 4 | Performance of HS image reconstruction for HS image datasets (CAVE, Tokyo Tech and Manchester). a, Ground truth images, reconstructed images and absolute error maps of the reconstructed images showing the localization of the reconstruction error. The ground truth and reconstructed HS images are rendered as synthetic RGB images, showing no apparent difference, whereas the reconstruction error is localized at the boundary of structures where the spatial sparsity is low. Note that the overall error is still below 10%, which is comparable to a conventional computational imaging technique. Images adapted with permission from: CAVE, ref. [34], Columbia Imaging and Vision Laboratory; Tokyo Tech, ref. [35], IEEE; Manchester, ref. [36], The Optical Society. **b**. The improvement of PSNR as a function of the number of iterations of the image reconstruction algorithm reveals that a satisfactory image quality (PSNR > 30 dB) is obtained after 50 iteration steps. **c**, The dependence of the frame rate on the image resolution for the iterative image reconstruction with 50 iteration steps, demonstrating video-rate VGA HS imaging (see Supplementary Video 1). With the AI-based method, video-rate full-HD HS imaging is also achieved.

reconstructed HS images are rendered as synthetic RGB images, showing no apparent difference. For a quantitative analysis, the reconstruction errors are visualized as absolute error maps in the bottom row of Fig. 4a. The reconstruction error is localized at the edges of structures where the spatial sparsity is low (Supplementary Section 6); however, the absolute error is still below 10% at the edges of structures, which is comparable with previous reports[37,38]. Such an error at the edges can be further reduced by optimizing the reconstruction algorithm[39].

We also evaluated the convergence and the frame rate of the image reconstruction. Figure 4b shows the simulated image quality (that is, the peak signal-to-noise ratio, PSNR) as a function of the iteration step of the image reconstruction algorithm. The PSNR saturates to more than 30 dB after around 50 iteration steps for all datasets. These simulated PSNR values are comparable with those in past reports[30,31,37,38,40].

As our HS camera enables a fast shutter speed (~10 ms at 550 lux with f/4.0 aperture), the speed of the image reconstruction is slower than the shutter speed and limits the frame rate of our HS camera in the most cases. The frame rates (limited by the post-image reconstruction process, see Methods) were experimentally measured using an iterative method with 50 iteration steps. Figure 4c shows the frame rates as a function of the number of pixels, that is, the size of measurement matrix in equation (2), revealing an exponential dependency. With the commercially available single graphics processing unit (GPU) that we used here (see Methods), the frame rate reaches 32.3 fps for VGA resolution (640 × 480 pixels) and 7.14 fps at full-HD resolution (1,920 × 1,080 pixels) (see Supplementary Video 1). For further acceleration, non-iterative image reconstruction is also attempted with the help of AI, demonstrating video-rate (34.4 fps) operation at full-HD resolution.

In conclusion, we demonstrated a video-rate HS camera with RGB camera-comparable sensitivity and resolution. Our HS camera was fabricated by integrating a CMOS-compatible spatially and spectrally random coded mask onto a monochromatic image sensor. The measured filter transmittance was 45% for visible light, and the spatial resolution was 3 px for 3 dB contrast, thus making the camera comparable with a standard RGB camera. Hyperspectral images with an absolute error of 2.2% on average were obtained through iterative image reconstruction assuming spatial sparsity. A frame rate of 32.3 fps is achieved for VGA resolution by using iterative image reconstruction and further enhanced to full-HD resolution by using AI-based reconstruction. The advantages of our HS camera (that is, practical sensitivity, compact size and data compression) hold great promise for the adoption of HS imaging technologies in various scenarios, including consumer applications such as smartphones, drones and Internet of Things (IoT) devices.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41566-022-01141-5.

References

1. Lee, Z. et al. Model for the interpretation of hyperspectral remote-sensing reflectance. *Appl. Opt.* **33**, 5721–5732 (1994).
2. Lee, Z., Carder, K. L., Mobley, C. D., Steward, R. G. & Patch, J. S. Hyperspectral remote sensing for shallow waters. I. A semianalytical mode. *Appl. Opt.* **37**, 6329–6338 (1998).
3. Govender, M., Chetty, K. & Bulcock, H. A review of hyperspectral remote sensing and its application in vegetation and water resource studies. *Water SA* **33**, 145–151 (2007).
4. Mehl, P. M., Chen, Y.-R., Kim, M. S. & Chan, D. E. Development of hyperspectral imaging technique for the detection of apple surface defects and contaminations. *J. Food Eng.* **61**, 67–81 (2004).

5. Blackburn, G. A. Hyperspectral remote sensing of plant pigments. *J. Exp. Bot.* **58**, 855–867 (2006).
6. Dale, L. M. et al. Hyperspectral imaging applications in agriculture and agro-food product quality and safety control: a review. *Appl. Spectrosc. Rev.* **48**, 142–159 (2013).
7. Backman, V. et al. Detection of preinvasive cancer cells. *Nature* **406**, 35–36 (2000).
8. Bacon, C. P., Mattley, Y. & DeFrece, R. Miniature spectroscopic instrumentation: applications to biology and chemistry. *Rev. Sci. Instrum.* **75**, 1 (2004).
9. Lu, G. & Fei, B. Medical hyperspectral imaging: a review. *J. Biomed, Optics* **19**, 010901 (2014).
10. Jahr, W., Schmid, B., Schmied, C., Fahrbach, F. O. & Huisken, J. Hyperspectral light sheet microscopy. *Nat. Commun.* **6**, 7990 (2015).
11. Pian, Q., Yao, R., Sinsuebphon, N. & Intes, X. Compressive hyperspectral time-resolved wide-field fluorescence lifetime imaging. *Nat. Photon.* **11**, 411–414 (2017).
12. Cutrale, F. et al. Hyperspectral phasor analysis enables multiplexed 5D in vivo imaging. *Nat. Methods* **14**, 149–152 (2017).
13. Hadoux, X. et al. Non-invasive in vivo hyperspectral imaging of the retina for potential biomarker use in Alzheimer's disease. *Nat. Commun.* **10**, 4227 (2019).
14. Birk, R. J. & McCord, T. B. Airborne hyperspectral sensor systems. *IEEE Aerosp. Electron. Syst. Mag.* **9**, 26–33 (1994).
15. Brelstaff, G. J., Parraga, A., Troscianko, T. & Carr, D. Hyperspectral camera system: acquisition and analysis. In *Proc. SPIE Geographic Information Systems, Photogrammetry, and Geological/Geophysical Remote Sensing* (eds Lurie, J. B. et al.) Vol. 2587, 150–159 (SPIE, 1995).
16. Tack, N., Lambrechts, A., Soussan, P. & Haspeslagh, L. A compact, high-speed, and low-cost hyperspectral imager. *Proc. SPIE OPTO* **8266**, 82660Q (2012).
17. Lapray, P.-J., Wnag, X., Thomas, J.-B. & Gouton, P. Multispectral filter arrays: recent advances and practical implementation. *Sensors* **14**, 21626–21659 (2014).
18. Candes, E. J. & Wakin, M. B. An introduction to compressive sampling. *IEEE Signal Process Mag.* **25**, 21–30 (2008).
19. Bao, J. & Bawendi, M. G. A colloidal quantum dot spectrometer. *Nature* **523**, 67–70 (2015).
20. Sahoo, S. K., Tang, D. & Dang, C. Single-shot multispectral imaging with a monochromatic camera. *Optica* **4**, 1209–1213 (2017).
21. Wang, Z. et al. Single-shot on-chip spectral sensors based on photonic crystal slabs. *Nat. Commun.* **10**, 1020 (2019).
22. Oiknine, Y., August, I., Farber, V., Gedalin, D. & Stern, A. Compressive sensing hyperspectral imaging by spectral multiplexing with liquid crystal. *J. Imaging* **5**, 3 (2019).
23. Kurokawa, U., Choi, B. I., Chang & Chang, C.-C. Filter-based miniature spectrometers: spectrum reconstruction using adaptive regularization. *IEEE Sens. J.* **11**, 1556–1563 (2011).
24. Oliver, J., Lee, W., Park, S. & Lee, H.-N. Improving resolution of miniature spectrometers by exploiting sparse nature of signals. *Opt. Express* **20**, 2613–2625 (2012).
25. Oliver, J., Lee, W.-B. & Lee, H.-N. Filters with random transmittance for improving resolution in filter-array-based spectrometers. *Opt. Express* **21**, 3969–3989 (2013).
26. Wang, Z. & Yu, Z. Spectral analysis based on compressive sensing in nanophotonic structures. *Opt. Express* **22**, 25608–25614 (2014).
27. Gehm, M. E. et al. Static two-dimensional aperture coding for multimodal, multiplex spectroscopy. *Appl. Opt.* **45**, 2965–2974 (2006).
28. Gehm, M. E., John, R., Brady, D. J., Willett, R. M. & Schulz, T. J. Single-shot compressive spectral imaging with a dual-disperser architecture. *Opt. Express* **15**, 14013–14027 (2007).
29. Wagadarikar, A., John, R., Willett, R. & Brady, D. Single disperser design for coded aperture snapshot spectral imaging. *Opt. Express* **47**, B44–B51 (2008).
30. Kittle, D., Choi, K., Wagadarikar, A. & Brady, D. J. Multiframe image estimation for coded aperture snapshot spectral imagers. *Appl. Opt.* **49**, 6824–6833 (2010).
31. Arce, G. R., Brady, D. J., Carin, L., Arguello, H. & Kittle, D. S. Compressive coded aperture spectral imaging: an introduction. *IEEE Signal Process Mag.* **31**, 105–115 (2014).
32. Bioucas-Dias, J. E. & Figueiredo, M. A. T. A new TwIST: two-step iterative shrinkage/thresholding algorithms for image restoration. *IEEE Trans. Image Process.* **16**, 2992–3004 (2007).
33. Tang, J., Nett, B. E. & Chen, G.-H. Performance comparison between total variation (TV)-based compressed sensing and statistical iterative reconstruction algorithms. *Phys. Med. Biol.* **54**, 5781–5804 (2009).
34. Yasuma, F., Mitsunaga, T., Iso, D. & Nayar, S. K. Generalized assorted pixel camera: post-capture control of resolution, dynamic range and spectrum. *IEEE Trans. Image Processing* **19**, 2241–2253 (2008).
35. Monno, Y., Kikuchi, S., Tanaka, M. & Okutomi, M. A practical one-shot multispectral imaging system using a single image sensor. *IEEE Trans. Image Process.* **24**, 3048–3059 (2015).
36. Foster, D. H., Amano, K., Nascimento, S. M. C. & Faster, M. J. Frequency of metamerism in natural scenes. *J. Opt. Soc. Am. A* **23**, 2359–2372 (2006).
37. Zhang, T., Fu, Y., Wang, L. & Huang, H. Hyperspectral image reconstruction using deep external and internal learning. In *IEEE/CVF International Conference on Computer Vision* (IEEE, 2019).
38. Zhao, Y. et al. Hyperspectral imaging with random printed mask. In *IEEE Conference on Computer Vision and Pattern Recognition* 10149–10157 (IEEE, 2019).
39. Liming, T. & Zhuang, F. Edge and contrast preserving in total variation image denoising. *EURASIP J. Adv. Signal Process.* **2016**, 13 (2016).
40. Wang, P. & Menon, R. Computational multispectral video imaging. *J. Opt. Soc. Am. A* **35**, 189–199 (2018).

Methods

Design of the coded mask

The rigorous coupled-wave analysis (RSoft, DiffractMOD) was used to simulate the transmittance of the Fabry–Pérot filters with different cavity thicknesses. The refractive indices used in the simulation were determined by spectroscopic ellipsometry of the deposited films.

Fabrication of the coded mask

The fabrication process started with an ion-assisted physical vapour deposition of a bottom DBR and a cavity on a 15 mm diameter SiO_2 wafer with a thickness of 625 μm. The cavity layer was patterned using a mask with a random pattern at the pixel level by standard photolithography and etching process. By changing the random pattern and etching depth, the process was repeated, thereby forming 32 different cavity thicknesses randomly distributed over the mask. A top DBR was deposited onto the patterned cavity and then removed using a mask with a random pattern for half of the pixels.

Integration of the coded mask on the image sensor

Using optical adhesive, the coded mask was directly attached onto a monochromatic image sensor (ams, CMV2000-3E12M1PP), where the top DBR and the sensor surface were facing each other. The air gap was controlled by introducing 5-μm-thick epoxy resin (SU-8) pillars between the mask and the sensor surface. Due to the spatial randomness of the mask, the image quality was robust against the lateral misalignment, whereas rotation was controlled to be nearly zero to avoid undesirable moire interference between the pixel pitch of the mask and sensor. Large mismatch of the pixel pitch between the filters and the image sensor results in spatial averaging of the filter transmittance, thereby deteriorating the spatial randomness. For mass production, the coded mask was also monolithically integrated on an 8-inch diameter image sensor wafer by using a CMOS-compatible process at Imec (see ref. [16] for more detail). The preliminary result of our prototyping shows the same performance as this work, demonstrating an absolute average error of 2.6% over 20 bands. As the coded mask can be implemented by both on- and off-chip processes, our approach can be readily applicable to different types of sensors at various wavelengths.

Calibration of the camera

The measurement matrix of our HS camera was experimentally determined as follows; monochromatic light with a full-width at half-maximum of 10 nm was obtained from a supercontinuum laser (Leukos, Rock-450-5) with a monochromator (Shimadzu, SPG-120S-REV), and was introduced into an integrating sphere (Labsphere, HAS-08L). The spatially uniform monochromatic illumination from an exit of the integrating sphere was then captured by our HS camera. By changing the wavelength band, the transmittance patterns of the coded mask were measured to determine the measurement matrix in equations (1) and (2). To compensate optical system properties with the measurement matrix, the lens settings (focal length and F number) during the imaging should be the same as those in the calibration process. By preparing presets of the measurement matrix associated with the different calibration settings, the lens settings and the wavelength bands can be changed in the post-image reconstruction process.

Characterization of the camera

An RGB camera (Ximea, MQ022CG-CM) with an equivalent sensor and a Bayer filter mosaic (ams, CMV2000-3E5C1PP) was used for fair comparison of the sensitivity and resolution. The objects in Fig. 3a,b were illuminated by standard LED lighting with the illuminance of 550 lux in total. To capture the objects, the exposure time, digital gain, analogue gain and gamma were set at 20 ms, 3.2 dB, 1 dB and 1.0, respectively. The ladder resolution chart printed on a white paper and the color samples in Fig. 3c,d,f were illuminated by a pair of wideband LED lighting (effilux, EFFI-FLEX-HSI) with the illuminance of 550 lux in total. To capture the resolution chart, the exposure time, digital gain, analogue gain and gamma were set to 5 ms, 3.2 dB, 1 dB and 1.0, respectively. To capture the colour samples on a black paper, the exposure time, digital gain, analogue gain and gamma were set to 20 ms, 3.2 dB, 1 dB and 1.0, respectively. For the resolution chart and MTF curves, the HS image at $\lambda = 550$ nm was compared with the green component of the RGB camera. The spectra of the colour samples were obtained from the HS image by averaging 50×50 pixels at the center of each sample. For the ground truth, a spectral-scanning HS camera was assembled using bandpass filters (Thorlabs, FB'XXX'-10, where 'XXX' is from 450 nm to 650 nm in 10 nm steps) and a monochromatic camera with the same image sensor (ams, CMV2000-3E12M1PP). A 16-mm fixed focus objective lens (Edmund optics, no. 59-870) was used with f/4.0 aperture in all scenes.

Iterative/AI-based image reconstruction

Iterative image reconstruction was developed based on the TwIST algorithm[32], but we modified and optimized the algorithm for a much faster convergence rate in a GPU environment (NVIDIA, GeForce RTX 2080Ti); TV was used for the regularization term with the coefficient τ of 0.5.

AI-based image reconstruction was developed using a convolutional neural network-based method[38]. To further improve the frame rate, the size of convolutional neural network model was effectively reduced with maintaining the image reconstruction quality, followed by model optimization using NVIDIA TensorRT. With reduction of the number of parameters from 31 million to 2 million, the AI-based method exhibits 206.6 fps for VGA resolution (640×480 pixels), 76.9 fps at HD resolution ($1,280 \times 720$ pixels) and 34.4 fps at full-HD resolution ($1,920 \times 1,080$ pixels) in the same GPU environment as the iterative-based method. The image quality of the AI-based method was also evaluated using CAVE database[34], where the HS image sets were divided into training and test groups with a ratio of 80:20. Under this training condition, the AI-based method shows PSNR of 29.21 dB on average, which is a similar level to previous attempts[30,31,37,38,40]. Additional training naturally improves the image quality, reaching the same level as that of the iterative-based method.

Data availability

The main data supporting the findings of this study are available in Supplementary Information. Further data that support the findings of this study are available from the corresponding authors on reasonable request.

Author contributions

A.I. designed and directed the project. M.Y. performed the device design, the numerical simulation, and the data analysis. Y.Y. performed the sample fabrication and C.H. performed the sample characterization. T.K. developed the iterative image reconstruction, and A.N. developed the AI-based method. K.T. and N.S. supervised the development of the CMOS-compatible fabrication process for mass production at Imec. T.H. supervised the whole project. The manuscript was mainly written by M.Y. with input from A.I., and reviewed by all of the authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41566-022-01141-5.

Correspondence and requests for materials should be addressed to Atsushi Ishikawa.

Peer review information *Nature Photonics* thanks the anonymous reviewers for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.