Mid-infrared snapshot spectral imaging via nonlinear radial dispersion - Supplementary Information -

Jianan Fang,[1,2] Kun Huang,[1,2,3,*] Ruiyang Qin,[1] Jixi Zhang,[1] and Heping Zeng[1,2,4,5,†]

[1] *State Key Laboratory of Precision Spectroscopy, and Hainan Institute, East China Normal University, Shanghai 200062, China*

[2] *Chongqing Key Laboratory of Precision Optics, Chongqing Institute of East China Normal University, Chongqing 401121, China*

[3] *Collaborative Innovation Center of Extreme Optics, Shanxi University, Taiyuan, Shanxi 030006, China*

[4] *Shanghai Research Center for Quantum Sciences, Shanghai 201315, China*

[5] *Chongqing Institute for Brain and Intelligence, Guangyang Bay Laboratory, Chongqing, 400064, China*

Supplementary Note 1: Theoretical model for compressive spectral imaging

Coded aperture snapshot spectral imaging (CASSI) is a representative compressive spectral imaging architecture that enables acquisition of a three-dimensional spatio-spectral datacube from a single two-dimensional measurement. For a target scene with spectral radiance $f_s(x, y, \lambda)$, the CASSI system first applies spatial encoding using a coded aperture with transmittance $T(x, y)$. The encoded spectral image is then sheared by a wavelength-dependent dispersive element, typically a grating or prism, and finally integrated by a two-dimensional detector array in a single exposure.

Assuming spectral dispersion along the x-direction, the spectral density incident on the detector can be written as [1]

$$\begin{aligned} f_i(x, y, \lambda) = \iint & \delta(x' - [x + \alpha(\lambda - \lambda_c)])\delta(y' - y) \\ & f_s(x', y', \lambda)T(x', y')\eta(\lambda)\ \mathrm{d}x'\mathrm{d}y', \end{aligned} \tag{1}$$

where α is the linear dispersion coefficient of the dispersive element, λ_c is the reference wavelength, and $\eta(\lambda)$ accounts for the wavelength-dependent system transmission and diffraction efficiency. After spectral integration by the detector, the measured intensity becomes

$$\begin{aligned} g(x, y) &= \int f_i(x, y, \lambda)\ \mathrm{d}\lambda \\ &= \int f_s(x + \alpha(\lambda - \lambda_c), y, \lambda)T(x + \alpha(\lambda - \lambda_c), y)\eta(\lambda)\ \mathrm{d}\lambda. \end{aligned} \tag{2}$$

In the present snapshot spectral imaging system, spectral dispersion is not introduced by an external linear optical element. Instead, wavelength-dependent spatial mapping arises intrinsically from the phase-matching conditions of nonlinear frequency upconversion, as illustrated in Fig. S1. Mid-infrared (MIR) spectral upconversion is implemented in a 4f imaging configuration with

*Electronic address: khuang@lps.ecnu.edu.cn
†Electronic address: hpzeng@phy.ecnu.edu.cn

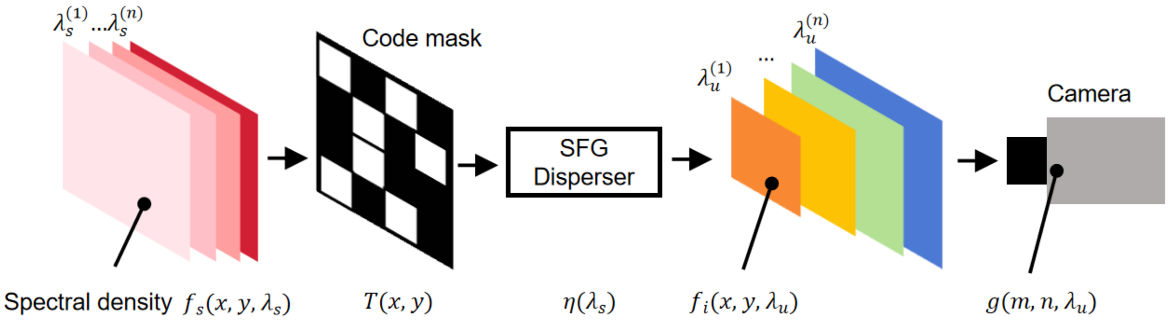

Figure S1: Schematic illustration of snapshot spectral encoding via intrinsic nonlinear radial dispersion. The scene is first spatially encoded by a mask (or diffuser-induced speckle), after which nonlinear frequency upconversion simultaneously performs wavelength conversion and wavelength-dependent radial dispersion. The resulting spectrally multiplexed image is recorded in a single exposure by a two-dimensional detector, enabling snapshot acquisition of spatio-spectral information without external dispersive optics.

a chirped-poling lithium niobate (CPLN) crystal placed at the Fourier plane. The nonlinear interaction is governed by sum-frequency generation (SFG), which satisfies the energy conservation relation

$$\frac{1}{\lambda_u} = \frac{1}{\lambda_s} + \frac{1}{\lambda_p}, \tag{3}$$

where λ_s, λ_p, and λ_u denote the vacuum wavelengths of the MIR signal, pump, and upconverted fields, respectively. Additionally, momentum conservation must be satisfied to achieve efficient frequency conversion. The phase-mismatch vector can be expressed in terms of axial and transverse components as

$$\frac{\Delta k_{\parallel}}{2\pi} = \frac{n_u}{\lambda_u}\cos\theta_u - \frac{n_s}{\lambda_s}\cos\theta_s - \frac{n_p}{\lambda_p} - \frac{1}{\Lambda} \ , \tag{4}$$

$$\frac{\Delta k_{\perp}}{2\pi} = \frac{n_u}{\lambda_u}\sin\theta_u - \frac{n_s}{\lambda_s}\sin\theta_s \ , \tag{5}$$

where $n_{s,p,u}$ are the refractive indices and $\theta_{s,p,u}$ denote the internal propagation angles inside the nonlinear crystal. By employing a chirped poling period Λ, the phase-matching condition $|\Delta\vec{k}| = \sqrt{\Delta k_{\parallel}^2 + \Delta k_{\perp}^2} = 0$ can be satisfied simultaneously for a broad range of signal wavelengths and incident angles. This broadband, wide-field phase-matching capability fundamentally differs from the narrowband response of uniformly poled lithium niobate crystals [2, 3], and forms the physical basis for snapshot MIR spectral imaging without wavelength or angle scanning [4].

From transverse momentum conservation, the external propagation angles of the signal and upconverted fields satisfy

$$\frac{\sin\theta'_u}{\sin\theta'_s} = \frac{\lambda_u}{\lambda_s} \ , \tag{6}$$

where $\theta'_{s,u}$ denote the angles outside the crystal and are related to the internal angles through Snell's law. In a 4f imaging system with relay lenses of focal lengths f_1 and f_2, this angular relation leads to a wavelength-dependent spatial magnification given by [5]

$$M(\lambda_s, \theta'_s) = \frac{f_2 \times \tan\theta'_u}{f_1 \times \tan\theta'_s} = \frac{f_2 \lambda_u}{f_1 \lambda_s} \times \sqrt{\frac{1 - \sin^2\theta'_s}{1 - \sin^2\theta'_s \times \lambda_u^2/\lambda_s^2}} \ . \tag{7}$$

Under the small-angle approximation, this expression simplifies to

$$M(\lambda_s) = \frac{f_2}{f_1} \times \frac{\lambda_u}{\lambda_s} = \frac{f_2}{f_1} \times \frac{1}{\lambda_s/\lambda_p + 1} \ , \tag{8}$$

revealing that each MIR wavelength is mapped to a distinct spatial magnification. This wavelength-dependent magnification produces a radially dispersive scaling of the image, constituting intrinsic nonlinear radial dispersion rather than linear translational shear.

Accordingly, the spatial coordinates transform as $x' = (x - x_0)/M(\lambda_s) + x_0$ and $y' = (y - y_0)/M(\lambda_s) + y_0$, where (x_0, y_0) denotes the center of radial scaling. After spectral integration, the snapshot measurement can be written as

$$\begin{aligned} g(x,y) &= \int f_i(x,y,\lambda_u)\ \mathrm{d}\lambda_u \\ &= \int f_s(\frac{x-x_0}{M(\lambda_s)} + x_0, \frac{y-y_0}{M(\lambda_s)} + y_0, \lambda_s)\eta(\lambda_s) \\ &\quad \cdot T(\frac{x-x_0}{M(\lambda_s)} + x_0, \frac{y-y_0}{M(\lambda_s)} + y_0)\ \mathrm{d}\lambda_s. \end{aligned} \tag{9}$$

where $\eta(\lambda_s)$ also includes the wavelength-dependent nonlinear conversion efficiency.

For discretization, let Δ denote the detector pixel pitch and assume a coded aperture pixel size $\Delta_{\text{code}} = q\Delta$ with $q \geq 1$. The encoding pattern can be expressed as

$$T(x,y) = \sum_{m,n} T(m,n)\mathrm{rect}\left(\frac{x}{q\Delta} - m, \frac{y}{q\Delta} - n\right) \ . \tag{10}$$

Integrating the continuous measurement over each detector pixel (m,n) yields

$$\begin{aligned} g(m,n) &= \int\int g(x,y)\mathrm{rect}\left(\frac{x}{\Delta} - m, \frac{y}{\Delta} - n\right)\mathrm{d}x\mathrm{d}y \\ &= \int\int \mathrm{d}x\mathrm{d}y\ \mathrm{rect}\left(\frac{x}{\Delta} - m, \frac{y}{\Delta} - n\right) \int f_s\left(\frac{x-x_0}{M(\lambda_s)} + x_0, \frac{y-y_0}{M(\lambda_s)} + y_0, \lambda_s\right)\eta(\lambda_s) \\ &\quad \times \left[\sum_{m',n'} T\left(\frac{\frac{x-x_0}{M(\lambda_s)} + x_0}{q\Delta} - m', \frac{\frac{y-y_0}{M(\lambda_s)} + y_0}{q\Delta} - n'\right)\right] \mathrm{d}\lambda_s \ . \end{aligned} \tag{11}$$

To enable computational reconstruction, the measurement and spectral datacube are vectorized as shown in Fig. S2. More specifically, we have

$$\begin{aligned} \boldsymbol{y} &= \mathrm{vect}[g(m,n)] \ , \\ \boldsymbol{x} &= \mathrm{vect}[f_s(m,n,\lambda_s)] \ , \end{aligned} \tag{12}$$

where the measurement term $g \in \mathbb{R}^{M\times N}$ and spectral cube $f_s \in \mathbb{R}^{M\times N\times L}$, with spatial dimension $M \times N$ and spectral dimension L. After vectorization, we have the vectorized terms $\boldsymbol{y} \in \mathbb{R}^{MN}$, $\boldsymbol{x} \in$

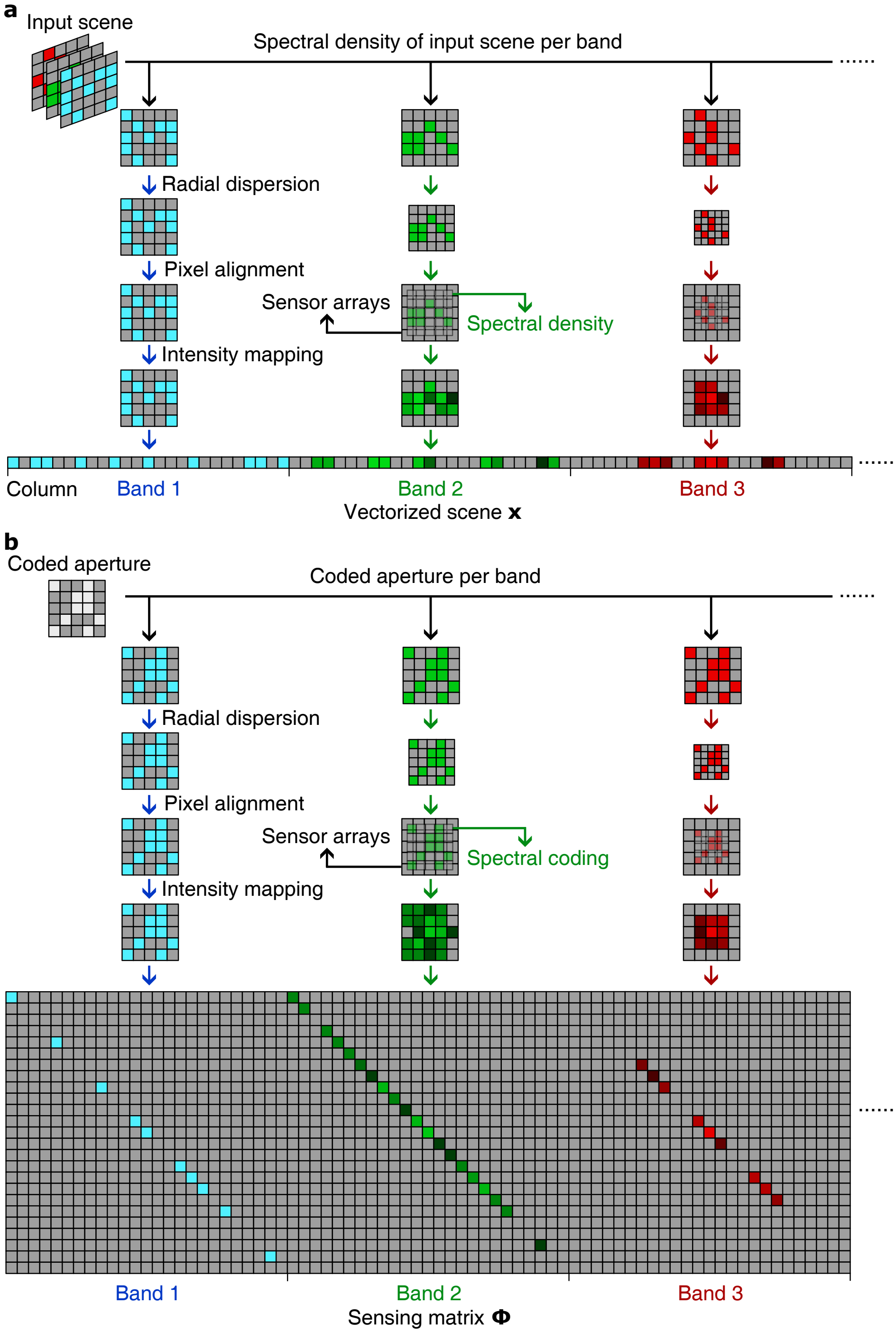

Figure S2: Vectorization of the spectral datacube and construction of the sensing matrix for nonlinear radial dispersion-based encoding. (a) Illustration of the vectorization process, in which a three-dimensional spectral datacube $f(m, n, l)$ is reshaped by stacking the vectorized two-dimensional spatial slices along the spectral dimension. (b) Schematic of sensing-matrix generation from the coded aperture in the nonlinear radial dispersion architecture. The sensing matrix is composed of wavelength-dependent diagonal blocks, each derived from the vectorized coded-aperture pattern of an individual spectral band and repeated along the spatial dimension according to the corresponding radial scaling.

$\mathbb{R}^{MNL}$. The combined effects of speckle encoding and wavelength-dependent radial dispersion are incorporated into a sensing matrix $\mathbf{\Phi} \in \mathbb{R}^{MN \times MNL}$, leading to the linear forward model

$$\boldsymbol{y} = \mathbf{\Phi}\boldsymbol{x} \ . \tag{13}$$

Reconstruction of the spectral datacube is then formulated as a regularized inverse problem,

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \frac{1}{2}\|\boldsymbol{y} - \mathbf{\Phi}\boldsymbol{x}\|_2^2 + \rho R(\boldsymbol{x}) \ , \tag{14}$$

where $R(\boldsymbol{x})$ encodes prior information about the scene, such as sparsity or total variation, and ρ balances data fidelity and regularization.

Supplementary Note 2: Algorithm for spectral imaging reconstruction

Reconstruction of the snapshot spectral datacube is formulated as a linear inverse problem constrained by prior knowledge of the underlying scene. Because the number of measurements is significantly smaller than the number of unknowns, the system is inherently underdetermined, and infinitely many solutions can satisfy the forward model. To obtain a physically meaningful and stable reconstruction, additional regularization must be imposed to restrict the solution space. In this work, we adopt the generalized alternating projection with total variation regularization (GAP-TV) algorithm to reconstruct the spectral datacube. Total variation (TV) regularization is well suited for spectral imaging problems, as it promotes piecewise-smooth spatial structures while preserving sharp edges. Compared with algorithms based on the alternating direction method of multipliers (ADMM), the GAP framework offers reduced parameter sensitivity and lower computational complexity, making it particularly attractive for large-scale compressive sensing problems such as snapshot spectral imaging [6].

Let $\boldsymbol{x}$ denote the vectorized spectral datacube and $\boldsymbol{y}$ the corresponding measurement vector. The reconstruction problem can be written as a TV-constrained feasibility problem,

$$\text{Find} \quad \min_{\boldsymbol{x},C} \quad \text{subject to} \quad \|\text{TV}(\boldsymbol{x})\| \leq C, \quad \mathbf{\Phi}\boldsymbol{x} = \boldsymbol{y}. \tag{15}$$

where C specifies the radius of the feasible TV ball and $\mathbf{\Phi}$ is the sensing matrix incorporating both spatial encoding and wavelength-dependent radial dispersion. The total variation operator is expressed as $\|\text{TV}(\boldsymbol{x})\| = \|\mathbf{D}\boldsymbol{x}\|_1$, where $\mathbf{D}$ denotes the discrete differential operator applied along the appropriate spatial and/or spectral dimensions of the datacube.

Within the GAP framework, the constrained problem in Eq. (S15) is solved via alternating Euclidean projections. Specifically, at iteration t, the reconstruction is obtained by solving

$$(\boldsymbol{x}^{(t)}, \boldsymbol{\theta}^{(t)}) \ = \ \arg\min_{\boldsymbol{x},\boldsymbol{\theta}} \ \tfrac{1}{2}\|\boldsymbol{x} - \boldsymbol{\theta}\|_2^2 \quad \text{subject to} \quad \|\text{TV}(\boldsymbol{\theta})\| \leq C^{(t)}, \ \mathbf{\Phi}\boldsymbol{x} = \boldsymbol{y} \ , \tag{16}$$

which is equivalent to the penalized formulation

$$(\boldsymbol{x}^{(t)}, \boldsymbol{\theta}^{(t)}) \ = \ \arg\min_{\boldsymbol{x},\boldsymbol{\theta}} \ \tfrac{1}{2}\|\boldsymbol{x} - \boldsymbol{\theta}\|_2^2 + \rho \, \|\text{TV}(\boldsymbol{\theta})\| \quad \text{subject to} \quad \mathbf{\Phi}\boldsymbol{x} = \boldsymbol{y} \ , \tag{17}$$

where ρ is a regularization parameter and t denotes the iteration index.

The optimization is carried out by alternately updating $\boldsymbol{x}$ and $\boldsymbol{\theta}$. With $\boldsymbol{\theta}$ fixed, the update of $\boldsymbol{x}$ corresponds to the Euclidean projection of $\boldsymbol{\theta}$ onto the affine constraint defined by the measurement equation, yielding

$$\boldsymbol{x}^{(t)} \;=\; \boldsymbol{\theta}^{(t-1)} + \boldsymbol{\Phi}^\top \big(\boldsymbol{\Phi}\boldsymbol{\Phi}^\top\big)^{-1} \big(\boldsymbol{y} - \boldsymbol{\Phi}\, \boldsymbol{\theta}^{(t-1)}\big), \tag{18}$$

where $\boldsymbol{\Phi}\boldsymbol{\Phi}^\top$ is assumed to be invertible. In many compressive sensing implementations, this matrix is diagonal or block-diagonal, allowing efficient computation of the inverse.

With $\boldsymbol{x}$ fixed, the update of $\boldsymbol{\theta}$ reduces to a TV denoising subproblem. This step is efficiently solved using an iterative clipping scheme,

$$\boldsymbol{\theta}^{(t)} = \boldsymbol{x}^{(t)} - \mathbf{D}^\top z^{(t)}, \tag{19}$$

$$\boldsymbol{z}^{(t)} = \mathrm{clip}\bigg(\boldsymbol{z}^{(t-1)} + \frac{1}{\alpha}\, \mathbf{D}\, \boldsymbol{\theta}^{(t-1)},\ \frac{\rho}{2}\bigg)\ , \tag{20}$$

where $\boldsymbol{z}$ is an auxiliary variable initialized as $\boldsymbol{z}^{(0)} = 0$, and α is chosen to satisfy $\alpha \geq \max\big(\mathrm{eig}(\mathbf{D}\mathbf{D}^\top)\big)$. The clipping operator is defined elementwise as

$$\mathrm{clip}(b, T) = \begin{cases} b, & \text{if } |b| \leq T, \\ T\ \mathrm{sign}(b), & \text{otherwise}\ . \end{cases} \tag{21}$$

Starting from $\boldsymbol{z}^{(0)} = 0$, the algorithm alternates between TV denoising and data-consistency projection until convergence. In practice, this procedure yields stable and accurate reconstructions for the snapshot spectral imaging measurements presented in this work.

Several factors can degrade the reconstruction fidelity of our current algorithm. First, the GAP-TV reconstruction assumes that the spatial distribution of the scene is piecewise smooth with sparse gradients. This assumption holds well for targets with distinct spatial features (e.g., letters, resolution targets, and patterned thin films), but may lead to reduced fidelity for highly textured or random scenes (e.g., biological tissues with fine structures or rough surfaces). Second, the spectral resolution is primarily limited by the calibration channel spacing (approximately 50 cm^{-1} in our implementation). Spectral features narrower than this spacing may not be fully resolved and could be smeared across adjacent channels.

To address these limitations, we envision several strategies. For highly textured scenes, alternative regularization schemes (e.g., wavelet-based sparsity or learned priors from deep generative models) could replace or complement total variation. For sub-calibration spectral features, a finer calibration step (e.g., using an AOTF with narrower bandwidth or a tunable laser source) could improve spectral resolution at the cost of increased calibration time. Alternatively, advanced super-resolution algorithms or dictionary learning methods could potentially extract sub-channel spectral information if the target spectra are known to lie within a low-dimensional manifold. These improvements are beyond the scope of the present work but represent promising directions for future research.

On a standard desktop computer (Intel Core i9-14900HK, 64 GB RAM, Python 3.9, without GPU acceleration), the GAP-TV algorithm with 100 iterations reconstructs a single 1024×1024×24 spectral datacube (24 spectral channels) in approximately 10 minutes. While the system acquires snapshots at 20 Hz, post-processing remains a bottleneck for real-time applications. However, significant acceleration is feasible: GPU acceleration offers substantial speedup, reducing reconstruction time to well under a minute; deep learning-based methods can achieve millisecond inference; and further algorithmic optimizations (sparse matrices, Numba JIT compilation) would provide additional gains. The reconstruction code is provided as open source to support such optimizations.

Supplementary Note 3: Pseudo-color rendering of simulated images

To validate the performance of the proposed snapshot spectral reconstruction scheme, numerical simulations are performed using the GAP-TV algorithm, with results presented in Fig. 2 of the main manuscript. The reconstructed images show close agreement with the corresponding ground truth. Here we describe the procedure used to generate the simulated spectral test images.

The base image used in the simulation is a grayscale image of a four-leaf clover, shown in Fig. S3(a). The image has a spatial resolution of 512×512 pixels with an 8-bit intensity depth. For the purpose of introducing spatially varying spectral signatures, the image is segmented into five regions corresponding to the four leaves and the stem. Spectral variation is synthetically assigned by modulating the grayscale intensity in each region with wavelength-dependent weighting coefficients. Specifically, for each wavelength, the original grayscale intensity in a given region is multiplied by a predefined spectral weight associated with that region. The spectral weighting functions used for the five regions are shown in Fig. S3(b), thereby defining distinct spectral responses across the simulated object.

To facilitate visual inspection of spectral differences, a pseudo-colour rendering is applied. For each wavelength, the intensity at every pixel is further mapped onto three color channels (R, G

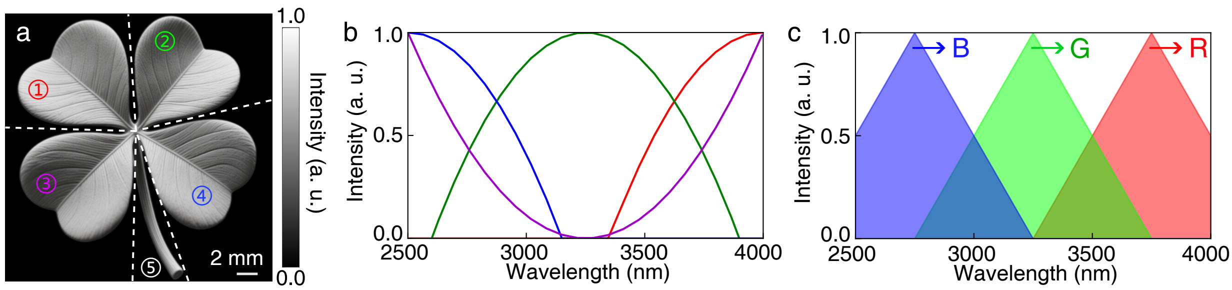

Figure S3: Pseudo-colour rendering procedure for the simulated spectral images. (a) Original grayscale image used in the numerical simulation, in which the four leaves and the stem of the clover are assigned distinct spectral signatures. For each wavelength, the pixel intensity is obtained by multiplying the original grayscale value by a region-specific, wavelength-dependent weighting function, whose spectral profiles are shown in (b). (c) The snapshot imaging spectral range (2.5-4.0 μm) is mapped onto the RGB color space using predefined channel-weighting functions, with the shaded regions indicating the relative contributions of the R, G, and B channels for visualization.

and B) using wavelength-dependent channel weighting coefficients. These RGB weighting curves are shown in Fig. S3(c). The final pixel color is obtained by combining the weighted RGB channels, enabling intuitive visualization of spectral contrast across the scene. Using this procedure, a synthetic hyperspectral image with spatially varying spectral signatures is constructed. The resulting pseudo-colour representation, in which different leaves of the clover appear in distinct colors, is shown in Fig. 2(A) of the main manuscript.

Supplementary Note 4: Details about experimental setup

A detailed schematic of the experimental system is shown in Fig. S4. The setup consists of three functional modules: (i) preparation and synchronization of the MIR signal and pump sources, (ii) broadband and wide-field nonlinear upconversion, and (iii) snapshot spectral imaging and detection. The MIR signal is provided by a supercontinuum (SC) fiber laser (Thorlabs, SC4500) covering a spectral range from 1.5 to 4.0 μm. The laser emits ultrashort pulses with a duration of approximately 300 fs at a repetition rate of 50 MHz. A long-pass filter with a cutoff wavelength of 2.4 μm (Edmund, #33-969) spectrally separates the output beam. The transmitted MIR component is further conditioned by a half-wave plate (Thorlabs, WPLH05M-2940) and a polarizer (Thorlabs, WP25M-IRA) to control the illumination power and enforce vertical polarization.

Prior to illuminating the target, the MIR beam is expanded to a diameter of 1 cm using a telescope formed by lenses L3 and L4, and then directed onto a reflective gold-coated diffuser (Thorlabs, DG10-1500-M01). The diffuser converts the incident MIR beam into spatially random speckle patterns, which act as a deterministic and calibratable spatial encoding of the object. By adjusting the separation between L3 and L4, the beam size on the diffuser is controlled, thereby enabling continuous tuning of the characteristic speckle size. The portion of the SC output reflected by the long-pass filter is coupled into an optical fiber, amplified to 15 mW, and injected into a nonlinear amplifying loop mirror (NALM) centered at 1.03 μm. Through cross-phase modulation inside the NALM cavity, passive synchronization between the MIR and pump pulses is achieved [7]. The cavity length of the NALM sets the repetition rate of the synchronized pump pulses to 25 MHz, such that half of the MIR pulses participate in the upconversion process. The pump pulse duration is approximately 10 ps, sufficient to temporally envelop the MIR pulses. Fine temporal overlap between the two pulse trains is electronically adjusted with a timing precision of 1 ps. After two stages of ytterbium-doped fiber amplification, the pump reaches an average power of 600 mW with a spectral bandwidth of $\sim$1 nm, ensuring high spectral fidelity during nonlinear conversion.

After interaction with the sample, the encoded MIR image is relayed by a $4f$ imaging system composed of two lenses with focal lengths $f_1 = 50$ mm and $f_2 = 75$ mm. A chirped-poling lithium niobate (CPLN) crystal is placed at the Fourier plane to enable sum-frequency generation. The crystal has dimensions of $3 \times 2 \times 10$ mm^3 (width$\times$thickness$\times$length), with a linearly varying poling period from 16 to 24 μm along the propagation direction. This chirped-poling design provides a large angular acceptance (up to $\sim$20°) over a broad spectral window from 2.4 to 5 μm, enabling wide-field and broadband nonlinear upconversion. The wavelength-dependent phase-matching condition of the CPLN crystal gives rise to intrinsic nonlinear radial dispersion, which

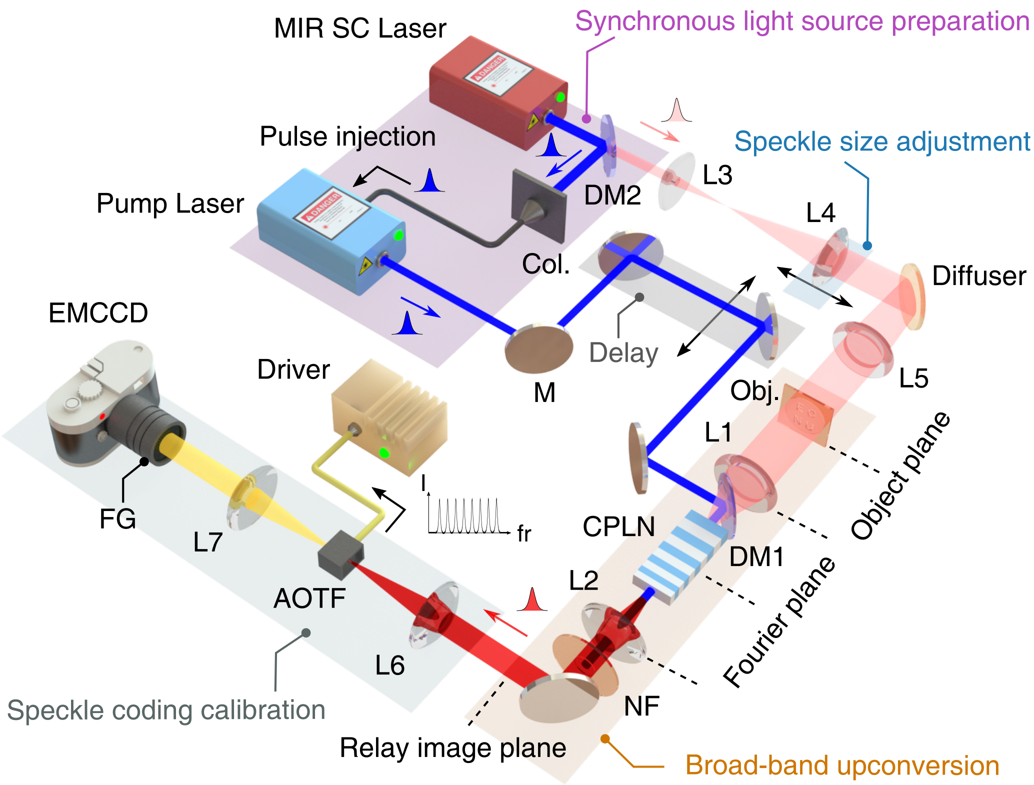

Figure S4: Schematic of the experimental setup for snapshot MIR spectral imaging based on nonlinear radial dispersion. The system comprises three functional modules: synchronous preparation of the MIR signal and pump sources, broadband and wide-field nonlinear upconversion, and snapshot spectral imaging and detection. CPLN, chirped-poling lithium niobate crystal; Col., collimator; NF, notch filter; DM, dichroic mirror; L, lens; M, silver mirror; Obj., object; AOTF, acousto-optic tunable filter; FG, filter group.

underpins the snapshot spectral encoding in a single exposure.

For calibration of the speckle encoding, the upconverted beam is directed to an acousto-optic tunable filter (AOTF) in the absence of an imaging target. The AOTF enables narrowband spectral selection with adjustable diffraction efficiency, allowing wavelength-resolved measurement of the speckle patterns. In addition, by applying composite radio-frequency signals, the AOTF supports parallel filtering and spectral flattening of the broadband upconverted signal, improving the dynamic range and conditioning of the subsequent reconstruction. The first-order diffracted beam from the AOTF is imaged onto a silicon-based EMCCD camera (Andor, iXon Ultra 888) with a sensor format of 1024×1024 pixels and a full-frame readout rate of up to 20 Hz, while the zero-order beam is blocked. To suppress background noise originating from pump-induced fluorescence and residual pump light, the upconverted signal passes through a cascade of interference filters, including a notch filter (Thorlabs, NF1064-44), a short-pass filter (Edmund, #64335), and a long-pass filter (Edmund, #68653). The overall transmission of the filtering stage is approximately 80%, while the suppression of the pump wavelength exceeds 180 dB.

The dark noise is an important figure of merit for sensitive optical imaging. In contrast to conventional MIR imagers based on HgCdTe and InSb, the silicon-based cameras are characterized by an extremely low dark current due to the much larger bandgap. The dark current of the EMCCD is specified to be about 8.7×10^{-4} electrons/pixel/second. In our experiment, the background noise of the upconversion imaging system is mainly caused by the pump-induced fluorescence. To characterize the imaging sensitivity, a noise equivalent power at the entrance of the MIR upconversion imager can be defined as

$$N_{noise} = \frac{N_p \times M}{\eta_{\mathrm{QE}} \times G} \times \frac{1}{\eta_{\mathrm{conv}} \times \eta_{\mathrm{filter}}}\ , \tag{22}$$

where $N_p = 3000$ counts/pixel/second is the pump-induced noise count, η_{QE}=80% is the quantum efficiency at 771 nm for the EMCCD camera, M=4.93 electrons/count is the A/D conversion factor, G=1000 is the gain factor, $\eta_{\mathrm{conv}}=10^{-3}$ is the conversion efficiency, and η_{filter}=80% is the total filtering efficiency. The given parameters result in $N_{\mathrm{noise}}=2.3 \times 10^4$ photons/pixel/second. This means that the sensitivity demonstrated in the experiment has not yet reached the pump-induced noise limit, and the photon-starved scenario can be addressed by further increasing the exposure time of acquisition.

Supplementary Note 5: Speckle coding preparation and characterization

Accurate reconstruction of spectral information from snapshot measurements requires prior calibration of the wavelength-dependent coding matrices. In the present system, this calibration is performed by spectrally resolving the broadband upconverted signal using an AOTF and recording the corresponding speckle patterns at discrete wavelengths. To characterize the spectral response of the AOTF, a superluminescent light-emitting diode (SLED) is employed as a continuous-wave broadband source. The central wavelength of the AOTF transmission band is observed to shift linearly toward shorter wavelengths with increasing radio-frequency (RF) driving frequency, as shown in Fig. S5(a). The spectral performance of the AOTF is primarily determined by the intrinsic properties and geometric dimensions of the acousto-optic crystal [8]. As shown in Fig. S5(b), the measured spectral resolution of the filter is approximately 2.9 nm. The relatively narrow bandwidth arises from the extended interaction length of the crystal, which enforces a more stringent phase-matching condition for photon-phonon coupling [9].

Using the calibrated RF-to-wavelength mapping, band-by-band filtering of the broadband upconverted signal is performed to record speckle encodings at each wavelength channel. By appropriately adjusting both the RF frequency and amplitude, the intensity response of the speckle patterns is equalized across wavelengths, ensuring uniform signal levels for subsequent reconstruction. Representative calibrated speckle patterns are shown in Fig. S5(d). As the wavelength increases, the upconverted image exhibits a gradual reduction in spatial extent, consistent with the wavelength-dependent radial scaling imposed by the nonlinear phase-matching condition.

To assess the suitability of the speckle encodings for snapshot spectral reconstruction, we compute the Pearson correlation coefficients (PCCs) among the 24 calibrated speckle coding matrices. The resulting correlation matrix is shown in Fig. S5(c). High PCC values indicate strong overlap and redundancy between codes, whereas low values correspond to approximately orthogonal and complementary encodings that are favorable for compressive inversion. As expected, higher

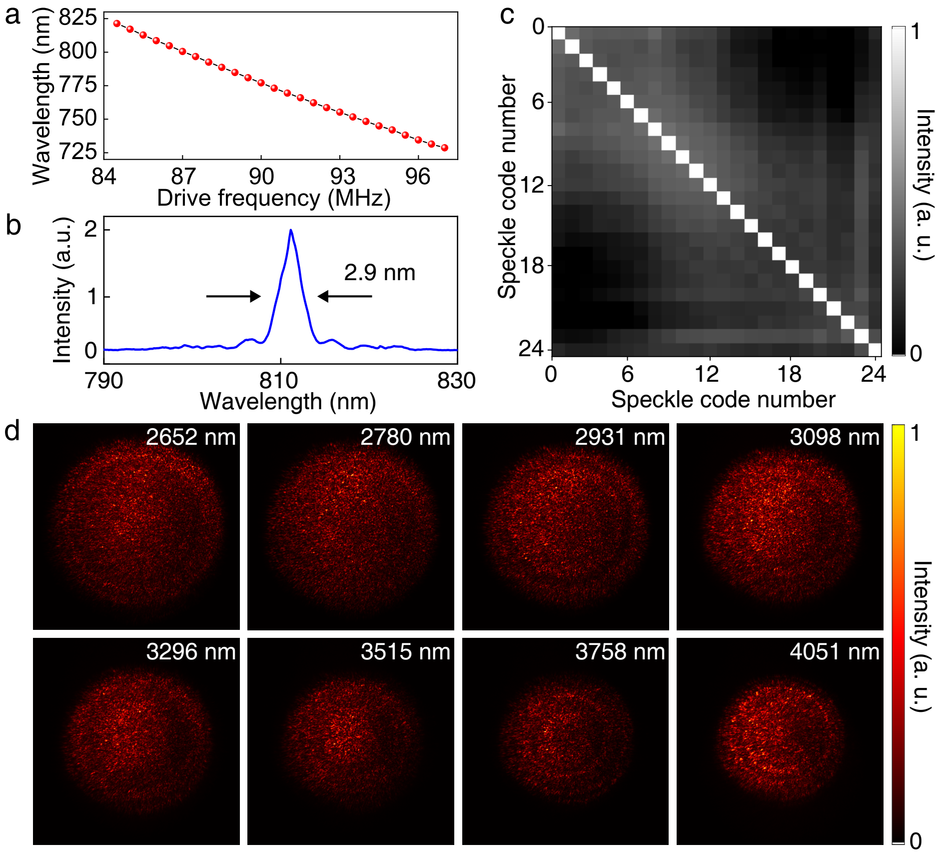

Figure S5: Calibration and characterization of speckle coding using an AOTF. (a) Measured dependence of the filter central wavelength on the applied radio-frequency (RF) driving frequency, exhibiting a linear tuning behavior. (b) Representative AOTF transmission spectra, showing a wavelength bandwidth of approximately 2.9 nm. (c) Pearson correlation coefficient matrix between the calibrated speckle coding channels. (d) Representative speckle-encoded images corresponding to selected wavelength channels.

correlations are observed between speckle patterns at adjacent wavelengths, owing to their similar radial scaling factors. Nevertheless, for a wavelength spacing of 62.5 nm, all the PCC values are maintained below 0.3, demonstrating sufficient code diversity for reliable snapshot spectral image reconstruction.

To further evaluate the effectiveness of speckle-based encoding, numerical simulations are performed and compared with conventional binary and grayscale encoding schemes. The original test image is the university emblem, shown in Fig. S6(a). The image is modulated using binary, grayscale, and speckle encodings [Figs. S6(b-d)], respectively, and the corresponding spectral images are spatially multiplexed according to the wavelength-dependent radial scaling factors. All encoding matrices have a spatial resolution of 512×512 pixels. The peak signal-to-noise ratios (PSNRs) of the reconstructed spectral images across wavelengths are summarized in Fig. S6(e). Binary encoding yields the highest reconstruction fidelity, while speckle encoding achieves PSNR values approximately 1 dB lower but consistently outperforms grayscale encoding. Representative reconstructed images for each encoding strategy are shown in Figs. S6(f-h).

The influence of encoding resolution is further examined by reducing the speckle encoding

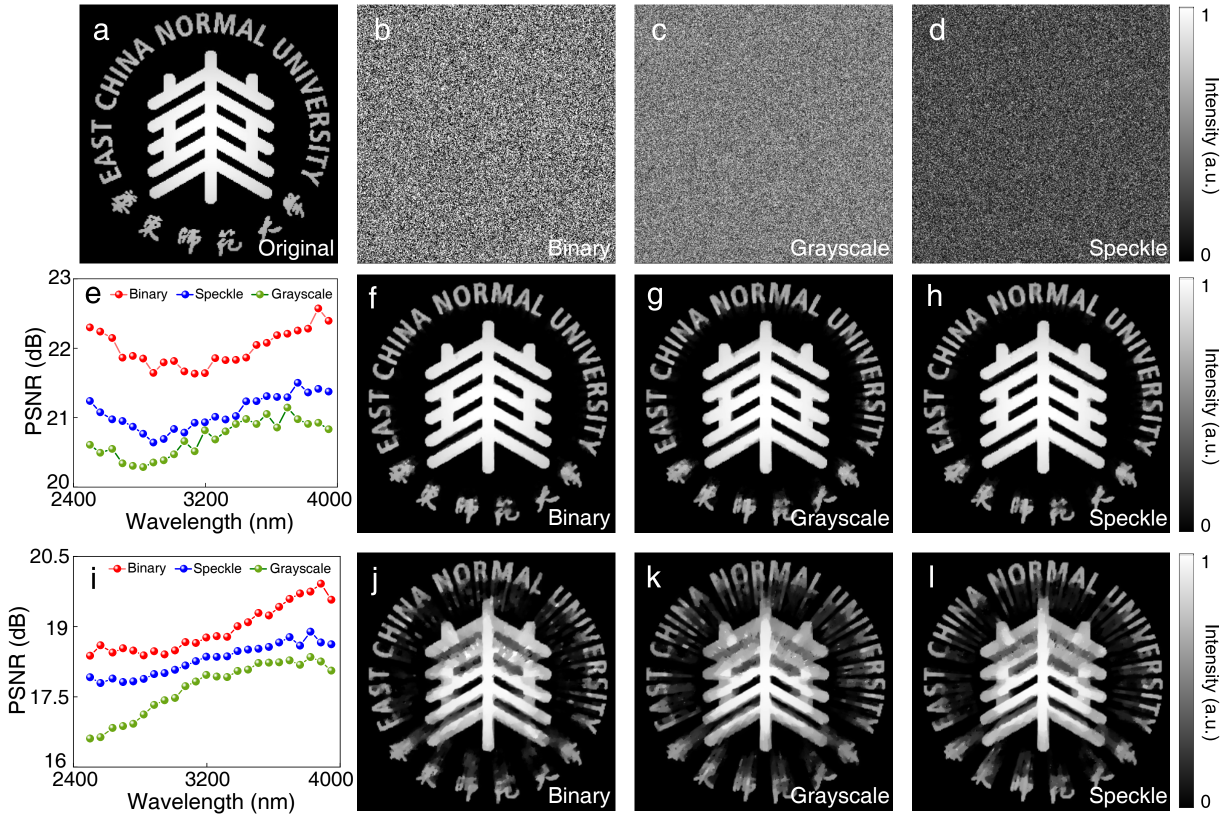

Figure S6: Numerical comparison of spectral image reconstruction performance using different encoding strategies. (a) Original target image used in the simulation (ECNU emblem). (b-d) Binary, grayscale, and speckle encoding patterns, respectively, each with a spatial resolution of 512×512 pixels. (e) Peak signal-to-noise ratio (PSNR) of the reconstructed spectral images as a function of wavelength for the three encoding schemes. (f-h) Representative reconstructed spectral images corresponding to the encoding patterns shown in (b-d). (i) PSNR of the reconstructed spectral images for different encoding schemes when the encoding resolution is reduced to 256×256 pixels. (j-l) Representative reconstructed spectral images obtained with different encoding schemes at a reduced encoding resolution of 256×256 pixels.

matrix size to 256×256. As illustrated in Fig. S6(i), this reduction leads to an approximately 3 dB decrease in PSNR compared with the higher-resolution case. Correspondingly, the reconstructed images exhibit increased blur and noise, as shown in Figs. S6(j-l). These results confirm that, while speckle-based encoding provides a practical and hardware-efficient alternative to binary masks, sufficient spatial resolution of the speckle patterns is critical to maintain high reconstruction fidelity.

Supplementary Note 6: Spectral resolution of the imaging system

The spectral resolution of the reconstructed snapshot images is jointly determined by several factors, including the bandwidth of the calibrated speckle encoding, the spectral blur introduced by the pump bandwidth during nonlinear upconversion, and the wavelength-dependent spatial separation imposed by nonlinear radial dispersion. Below we analyze these contributions and their combined effect on the achievable spectral resolution.

During calibration of the speckle encodings, an AOTF with a measured bandwidth of 2.9 nm is used to spectrally isolate the upconverted signal. In addition, the pump laser employed in the sum-frequency generation process has a spectral bandwidth of approximately 1 nm. Owing to energy conservation in nonlinear frequency conversion, these bandwidths introduce uncertainties in the inferred mid-infrared wavelength, which can be expressed as

$$\Delta\lambda_s^{(\mathrm{filter})} = \frac{\lambda_s^2}{\lambda_u^2}\Delta\lambda_u, \; \Delta\lambda_s^{(\mathrm{pump})} = \frac{\lambda_s^2}{\lambda_p^2}\Delta\lambda_p \; . \tag{23}$$

The corresponding values are plotted in Fig. S7(a). This analysis indicates that the dominant contribution arises from the calibration bandwidth of the speckle encoding, leading to an effective spectral uncertainty of approximately 50 nm over the operating wavelength range. In contrast, the narrow pump bandwidth does not constitute a primary limitation under the present conditions.

Beyond these bandwidth-related effects, spectral resolution in the snapshot configuration is fundamentally constrained by the spatial separability of wavelength-dependent image replicas produced by nonlinear radial dispersion. Unlike conventional gratings or prisms, which introduce translational dispersion independent of position, the radial dispersion in the present system depends both on wavelength and on the radial distance l from the scaling center. For two neighboring spectral components to be distinguishable, the radial displacement induced by their wavelength difference must exceed the characteristic size of the speckle encoding. This requirement can be expressed as

$$l|M(\lambda_c + \frac{1}{2}\Delta\lambda_s^{(\mathrm{scaling})}) - M(\lambda_c - \frac{1}{2}\Delta\lambda_s^{(\mathrm{scaling})})| \geq R_{\mathrm{speckle}}(\lambda_c)M(\lambda_c), \tag{24}$$

where λ_c is the center wavelength, $\Delta\lambda_s^{(\mathrm{scaling})}$ denotes the minimum resolvable wavelength separation imposed by radial scaling, and $R_{\mathrm{speckle}}(\lambda_c)$ is the speckle size at that wavelength. The

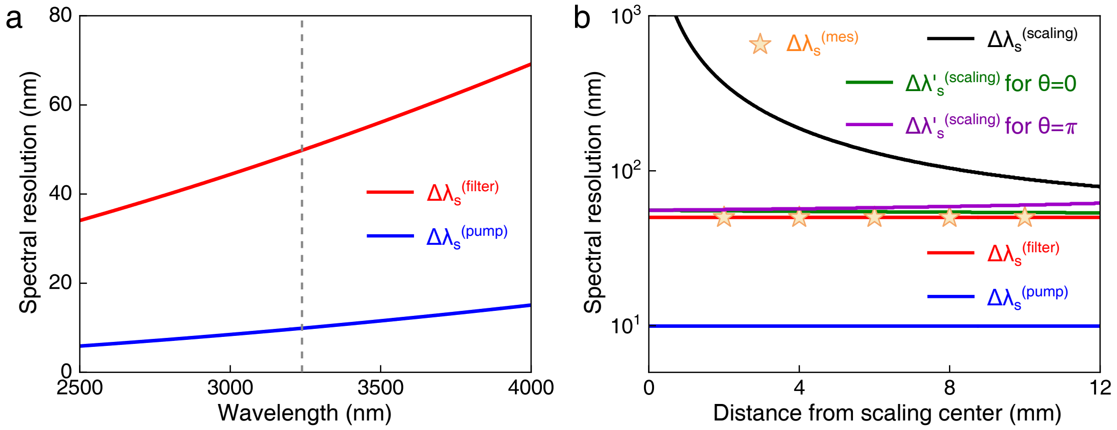

Figure S7: Theoretical analysis of the spectral resolution of the snapshot imaging system. (a) Calculated spectral uncertainty introduced by the calibrated speckle-encoding filter bandwidth and by the pump linewidth in the nonlinear upconversion process. (b) Spectral-resolution requirement imposed by wavelength-dependent nonlinear radial dispersion, as a function of the signal wavelength and the radial distance from the scaling center.

speckle size is extracted from the experimental measurements shown in Figs. 4(C, D) of the main manuscript.

When the equality in Eq. (S24) is satisfied, two spectral components become just distinguishable. Figure S7(b) plots the resulting spectral resolution limit imposed by radial dispersion for a representative center wavelength of $\lambda_c = 3250$ nm. As indicated by the black curve, the radial displacement decreases toward the scaling center, requiring a larger wavelength separation to achieve resolvability. Consequently, within approximately 1 mm of the scaling center, the effective spectral resolution is on the order of 1000 nm. At larger radial distances ($l > 8$ mm), the spectral resolution improves significantly to values below 100 nm.

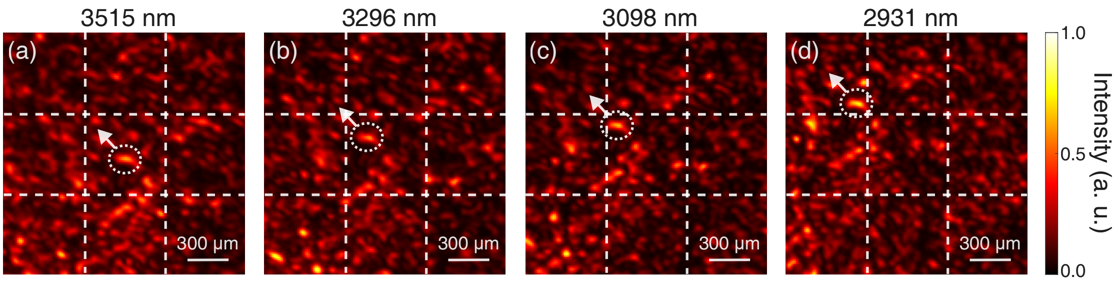

Figure S8: Wavelength-dependent translation dispersion induced by the diffuser. (a-d) show upconverted images of the speckle pattern at selected wavelengths (3515 nm, 3296 nm, 3098 nm, and 2931 nm). The lateral shift of characteristic speckle features across the field of view reveals the dispersion introduced by the diffuser. These shifts are used to determine the wavelength-dependent translation dispersion coefficients for calibration of the sensing matrix.

Experimentally, however, the measured spectral resolution, as extracted from a comparison between reconstructed spectra and reference FTIR measurements, shows relative insensitivity to the radial position. These experimentally obtained values, indicated by pentagram markers in Fig. S7(b), remain close to 50 nm across the field of view. This behavior arises from an additional translational dispersion introduced by the diffuser when reflecting the broadband MIR signal. Physically, this diffuser-induced translational dispersion originates from the ground-glass surface microstructure of the diffuser (Thorlabs DG10-1500-M01). The 1500-grit surface has micron-scale feature sizes that act as a grating. When the broadband MIR beam illuminates this surface at a 45° incident angle, different wavelengths are diffracted at slightly different angles, producing a wavelength-dependent spatial shift in the reflected beam. This diffuser-induced angular dispersion is calibrated into the sensing matrix $\mathbf{\Phi}$ alongside the radial dispersion from the nonlinear crystal, and together they provide sufficient wavelength-specific encoding diversity for the reconstruction algorithm to achieve uniform spectral resolution across the field of view. To this end, we selected calibration speckle images at four wavelengths, as shown in Figure S8. Based on the lateral shift of characteristic speckle features across the field of view, the corresponding translational dispersion coefficient is calculated to be $\alpha = 2.77$ μm/nm.

Accounting for this effect, the resolvability condition in Eq. (S24) must be modified to include both radial and translational contributions: $\sqrt{D_{\text{radial}}^2 + D_{\text{shift}}^2 + 2D_{\text{radial}}D_{\text{shift}}\cos\theta} \geq D_{\text{speckle}}$, where $D_{\text{shift}} = \alpha\,\Delta\lambda_s^{(\text{scaling})}$ denotes the translational displacement, and θ is the angle between the radial direction and the translational dispersion direction. Figure S7(b) shows the corrected spectral resolution $\Delta\lambda_s'^{(\text{scaling})}$ for θ=0 and $\theta = \pi$. Under these conditions, the spectral resolution varies smoothly with position and remains approximately 55 nm across the image.

Finally, we note that the spectral resolution of the snapshot imaging system can be further improved by optimizing several system parameters, including reducing the spatial size of the speckle encoding, employing calibration filters with narrower bandwidths, and using a pump source with reduced spectral linewidth.

Supplementary Note 7: Rapid identification of chemical films

The mid-infrared spectral region hosts fundamental molecular vibrational resonances that provide chemically specific contrast, making MIR spectroscopy a powerful tool for material identification and discrimination. The snapshot capability of our upconversion imaging system offers a potential solution for high-speed, non-destructive polymer discrimination. To validate this capability, we applied our system to discriminate between two different polymer thin-film materials based on their distinct MIR vibrational signatures.

As shown in Figure S9(a), we designed a target consisting of the letters "ECNU" engraved on an aluminum plate. The "EC" region was covered by a polystyrene (PS) thin film, while the "NU" region was covered by a polyethylene (PE) thin film. Both films had a thickness of approximately 20 μm. The target was illuminated by the broadband MIR supercontinuum source, and a single snapshot measurement was acquired using our upconversion imaging system with a 50 ms exposure time. Figure S9(b) shows the raw snapshot measurement acquired in a single exposure. From this single compressed measurement, we reconstructed the full spatio-spectral datacube using the GAP-TV algorithm described in Supplementary Note 2.

Figure S9(c) presents the reconstructed spectral responses of the PS film (red symbols) and the PE film (blue symbols), compared with reference FTIR spectra (solid lines) at the same resolution (50 cm^{-1}). The characteristic absorption features of each polymer are clearly resolved. Figure S9(d) shows reconstructed monochromatic images at different selected wavenumbers. These images reveal the spatial distribution of each polymer based on wavelength-dependent absorption contrast. At 3032 cm^{-1} (the PS absorption peak), the "EC" region (PS-covered) appears darker due to strong absorption, while the "NU" region (PE-covered) remains brighter. At 2836 cm^{-1} and 2908 cm^{-1} (PE absorption peaks), the opposite contrast is observed: the "NU" region appears darker, while the "EC" region remains brighter.

These results demonstrate that our snapshot MIR spectral imaging system can simultaneously acquire both spectral and spatial information from a single exposure, enabling rapid, non-destructive discrimination of different polymer materials. This capability suggests potential practical applications in areas such as rapid plastic sorting in recycling streams, quality control in polymer manufacturing, identification of plastics in environmental monitoring, and non-destructive testing of multilayer polymer films.

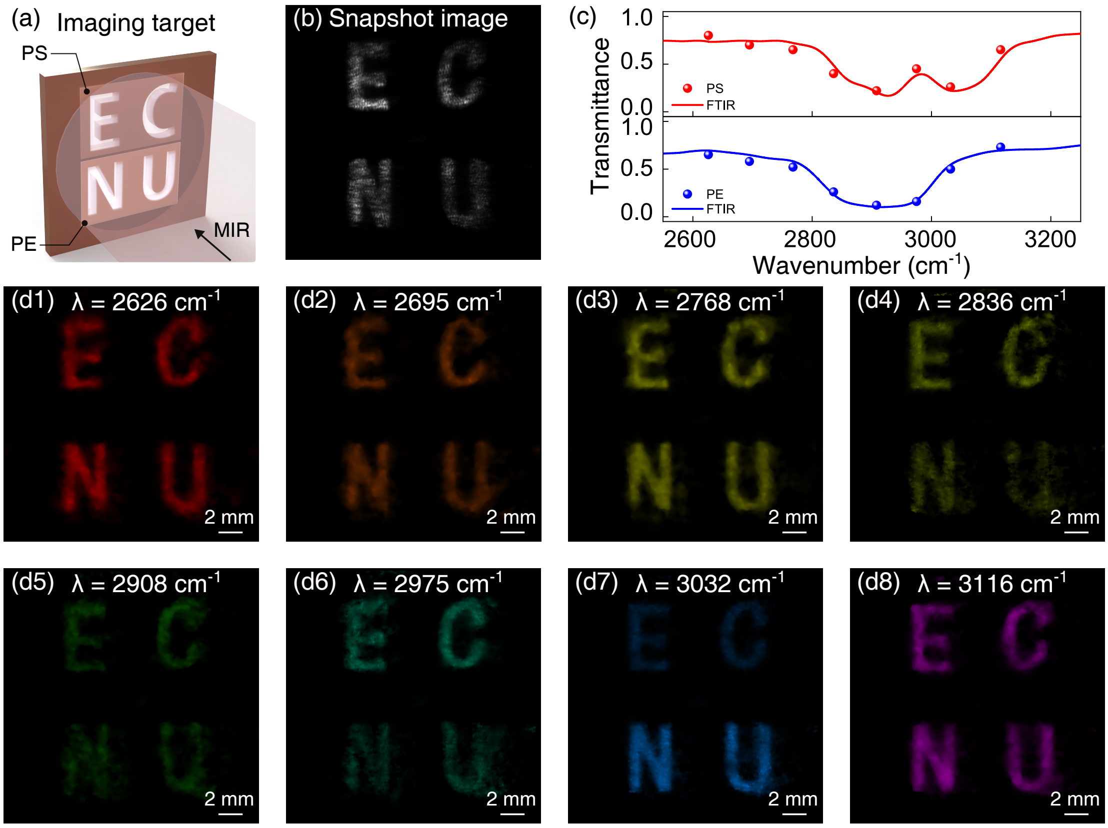

Figure S9: Chemically specific snapshot discrimination of polymer thin films. (a) The target comprises aluminum letters "ECNU" with "EC" covered by polystyrene (PS) and "NU" covered by polyethylene (PE). (b) Single-shot raw measurement. (c) Reconstructed spectra agree with FTIR references, showing distinct PS and PE vibrational fingerprints. (d) Reconstructed monochromatic images, demonstrating clear spatial-spectral contrast.

Supplementary Note 8: Comparison of wide-field MIR spectral imagers

To date, various approaches have been established for the implementation of wide-field MIR spectral imaging. A performance comparison of these methods is summarized in Table S1. The detection of mid-infrared signals typically relies on infrared focal plane arrays (FPA) based on microbolometers or narrow-bandgap semiconductors. For instance, the integration of a multi-pixel detector in Fourier transform infrared (FTIR) imaging spectrometers can significantly enhance data acquisition speeds [10, 11], offering a marked improvement over time-consuming point-scanning operations. However, the FTIR technique necessitates an interferometric scan followed by post-processing to reconstruct the MIR absorption spectrum, which can hinder the overall acquisition process. Alternatively, sequential band-by-band scanning of the target image is a common approach, often implemented by interchanging multiple bandpass filters. Recent advances in the electrical wavelength tuning of light sources and high-speed signal filtering have effectively accelerated information acquisition rates. Quantum cascade lasers (QCLs) have been widely employed for MIR spectral imaging due to their high spectral radiance, high spectral resolution, and rapid

Table S1: Performance comparison of wide-field MIR spectral imaging systems.

Type	Ref.	Source	Range (cm^{-1})	Res. (cm^{-1})	Band Num.	Pixel Num.	Time (per cube)	Power ($/nm/cm^2$)	Detector
SFG	This	SC	2500-4000	50	24	1024×1024	50 ms[a] 5 s[b]	0.8 μW 0.2 nW	Si
	[5]	SC	2600-4085	50	100	1024×1024	16 ms[c]	7 μW	Si
	[20]	OPO	2500-4348	3.6-7.6	62	658×496	25 s[d]	20 mW	Si
	[21]	DFG	640-3015	2.6-3.7	1069	640×480	8 s	1 μW	Si
NTA	[19]	OPO	2735-3265	8.4	63	1280×1024	1.1 s	17 μW	InGaAs
PTI	[16]	OPO	2830-2950	10	13	2016×2016	260 ms	4 W	Si
CS	[23]	Globar	2083-2703	10	111	640×512	20 ms	/[f]	InSb
	[22]	Globar	2000-3333	20	100	64×48	60 s	/[f]	InSb
Direct	[10]	Globar	988-1950	16	FTIR[e]	96×96	93 ms	/[f]	MCT
	[11]	Globar	1850-6667	1	FTIR[e]	256×160	60 s	/[f]	InSb
	[12]	QCL	952-1351	2	137	128×128	95 ms	18 μW	MCT
	[13]	QCL	777-1904	2	282	128×128	120 s	121 mW	MCT
	[14]	QCL	1160-1320	1	565	640×480	11.3 s	75 mW	Bolometer
	[15]	SC	2222-5000	1-5	100	640×512	1.2 s	7 mW	InSb

(a) Continuous shooting mode for dynamic scene capture at a frame rate of 20 Hz.

(b) Mid-infrared snapshot images collected under low illumination were recorded using an EMCCD set to -80 °C with a gain of 1000.

(c) The acquisition time is determined by the maximum frame rate of 6.4 kHz for the full spatial format of the CMOS camera used in our experiment.

(d) Although a wide-field image at a single wavelength can be recorded in 2.5 ms, the exposure time of the camera is set to be 400 ms for improving the image quality in the spectral imaging demonstration.

(e) The spectral imaging is realized based on the FTIR technique.

(f) The illumination power for the used thermal source is not given in these works.

wavelength tunability [12–14]. Furthermore, MIR spectral imaging spanning an octave can be realized by using multiple synchronized QCLs [13]. Recently, broadband supercontinuum (SC) laser sources have been leveraged to implement MIR hyperspectral imaging with the aid of an AOTF [15]. Nevertheless, in all the aforementioned direct approaches, performance is currently constrained by infrared imagers, which have long suffered from high dark noise, low pixel counts, and limited frame rates. Specifically, the frame rate for a 640×512 format is typically restricted to below 100 Hz [15]. While readout rates up to the kHz level are achievable, they come at the cost of a drastically reduced number of pixels, *e.g.*, a spatial format of 128×128 pixels [12]. Moreover, MIR imagers usually require cryogenic operation to suppress severe thermal noise and improve detection sensitivity.

To address these limitations of arrayed MIR detectors, indirect approaches have been developed wherein information encoded in the MIR region is transferred to the visible or near-infrared range to leverage mature large-bandgap semiconductor technology. This frequency conversion step can be realized through various mechanisms. For instance, MIR photothermal imaging (PTI) is implemented by using a visible laser to probe the infrared absorption-induced thermal lensing effect within the sample [16–18]. In wide-field PTI, the required intensive MIR illumination typically relies on an optical parametric oscillator (OPO) with high pulse energy. However, the spectral imaging speed is limited by the slow tuning speed of the OPO, particularly when covering a wide spectrum. Notably, the use of QCLs in MIR photothermal imaging provides access to fast-tuning illumination sources over a wide spectral range. When combined with a fast, high-sensitivity camera, the advent of high-power, high-repetition-rate QCLs could enable wide-field imaging modalities with significantly increased frame rates. Another attractive conversion method exploits the intrinsic optical nonlinearity of the detector material itself, based on non-degenerate two-photon absorption (NTA). Recently, NTA-based spectral imaging was implemented using a high-definition InGaAs camera [19]. While the spectral scanning rate can be fast during a single sweep, the dwell time between sequential sweeps is relatively long due to the mechanical inertia of the translational stage. Consequently, the demonstrated refresh rate for datacube acquisition is limited to the Hz level in continuous operation mode. It is worth mentioning that the refreshing rate in the NTA-based scheme could be substantially improved by using a voice coil actuator as the delay line, which would facilitate a real-time hyperspectral imaging at high definition.

Another conversion method exploits parametric interaction within a nonlinear crystal, which has been proven to be the most efficient method for spectral transduction based on the second-order SFG process [4]. In the presence of long nonlinear crystals, the phase-matching requirement generally leads to a narrow acceptance window for the infrared signal in both the spectral and spatial domains. Wide-field spectral upconversion imaging is typically realized by scanning phase-matching parameters, such as operation temperature or crystal orientation [20], which inevitably limits the spectral imaging speed. Very recently, broadband upconversion imaging was demonstrated using a thin crystal with a length shorter than 5 μm [21]. In that work, the broadband MIR source was prepared via four-wave difference frequency generation (DFG) based on a high-power Ti:Sapphire laser. Due to the extremely short interaction length, an ultra-high pump intensity provided by a complex solid-state laser system was required to obtain significant conversion efficiency. The involved high-energy pulses ($\sim$mJ) operated at a relatively low repetition rate ($\sim$ kHz), limiting the monochromatic imaging speed. Moreover, the filtering operation for the upconversion images was conducted by rotating a dispersive grating, imposing a stringent limitation on the spectral imaging speed. In our previous work, the combination of bright supercontinuum illumination, broadband frequency upconversion, fast acousto-optic filtering, and rapid wide-field detection enabled video-rate acquisition of mid-infrared spectral data cubes with large spatial formats and broad spectral channels. However, under photon-limited low-light conditions, the insufficient optical flux necessitates long camera integration times, and the sequential band-by-band scanning strategy loses the advantage of rapid filtering [5].

Compressed spectral imaging is a technique that exploits compressed sensing (CS) theory, utilizing encoding and reconstruction algorithms to achieve efficient spectral imaging. It reduces the data acquisition volume while maintaining high spatial, spectral, and temporal resolution; furthermore, the multiplexing of multispectral information in the image plane preserves photon throughput. Spatial encoding of the MIR field can be implemented with mid-infrared digital micromirror devices (MIR-DMD) equipped with custom windows, although such devices are relatively expensive [22]. A coded aperture snapshot spectral imaging system, combining a fixed coded mask with grating-based spectral dispersion, offers a more convenient and compact solution and has been shown to acquire 111 spectral bands in a single snapshot within 20 ms [23]. However, the large dark current noise of MIR focal plane arrays necessitates the use of high illumination power.

In this work, we design and implement ultrasensitive MIR snapshot spectral imaging at video frame rates based on a nonlinear radial dispersion process. In this scheme, the full MIR information in both spatial and spectral domains experiences simultaneous wavelength conversion and spatial dispersion during the upconversion process, while speckle illumination provides spatial encoding, eliminating the need for additional spatial modulation elements. Notably, controlling illumination intensity is critical in biochemical imaging to prevent sample damage. Here, the illumination intensity is approximately 0.2 $nW/nm/cm^2$, which is significantly lower than that of other spectral imaging systems. The illumination power could be further reduced by improving conversion efficiency. Therefore, the achieved features of high acquisition rates, wide-field operation, and broadband spectral coverage open up new possibilities for the high-throughput characterization of dynamic processes in chemical, medical, and bio-related fields.

[1] L. Huang, R. Luo, X. Liu and X. Hao, "Spectral imaging with deep learning," *Light Sci. Appl.* **11**, 61 (2022).

[2] Y. Wang, J. Fang, T. Zheng, Y. Liang, Q. Hao, E. Wu, M. Yan, K. Huang, and H. Zeng, "Mid-infrared single-photon edge enhanced imaging based on nonlinear vortex filtering," *Laser Photonics Rev.* **15**, 2100189 (2021).

[3] H. Zhang, S. Kumar, Y. M. Sua, S. Zhu, and Y.-P. Huang, "Near-infrared 3D imaging with upconversion detection," *Photonics Res.* **10**, 2760-2767 (2022).

[4] A. Barh, P. J. Rodrigo, L. Meng, C. Pedersen, and P. Tidemand-Lichtenberg, "Parametric upconversion imaging and its applications," *Adv. Opt. Photonics* **11**, 952-1019 (2019).

[5] J. Fang, K. Huang, R. Qin, Y. Liang, E Wu, M. Yan and H. Zeng, "Wide-field mid-infrared hyperspectral imaging beyond video rate," *Nat. Commun.* **15**, 1811 (2024).

[6] X. Yuan, "Generalized alternating projection based total variation minimization for compressive sensing," *Proc. IEEE Int. Conf. Inf. Process.*, 2539 (2016).

[7] J. Zeng, B. Li, Q. Hao, M. Yan, K. Huang and H. Zeng, "Passively synchronized dual-color mode-locked fiber lasers based on nonlinear amplifying loop mirrors," *Opt. Lett.* **44**, 5061-5064 (2019).

[8] Y. Pang, K. Zhang, and L. Lang, "Review of acousto-optic spectral systems and applications," *Front. Phys.* **10**, 1102996 (2022).

[9] L. Bei, G. I. Dennis, H. M. Miller, T. W. Spaine, and J. W. Carnahan, "Acousto-optic tunable filters: fundamentals and applications as applied to chemical analysis techniques," *Prog. in Quant. Electron.* **28**, 67-87 (2004).

[10] K. L. A. Chan, S. G. Kazarian, "FT-IR Spectroscopic Imaging of Reactions in Multiphase Flow in Microfluidic Channels," *Anal. Chem.* **84**, 4052-4056 (2012).

[11] J. Meléndez, G. Guarnizo, "Fast Quantification of Air Pollutants by Mid-Infrared Hyperspectral Imaging and Principal Component Analysis," *Sensors* **21**, 2092 (2021).

[12] D. B. Kelley, A. K. Goyal, N. Zhu, D. A. Wood, T. R. Myers, P. Kotidis, C. Murphy, C. Georgan, G. Raz, R. Maulini, A. Müller, "High-speed mid-infrared hyperspectral imaging using quantum cascade lasers," *Proc. SPIE* **10183**, 1018304 (2017).

[13] K. Yeh, S. Kenkel, J.-N. Liu, R. Bhargava, "Fast Infrared Chemical Imaging with a Quantum Cascade Laser," *Analytical chemistry* **87**, 485 - 493 (2015).

[14] K. Haase, N. Kröger-Lui, A. Pucci, A. Schönhals, W. Petrich, "Real-time mid-infrared imaging of living microorganisms." *J. Biophotonics* **9**, 61-66 (2016).

[15] M. Farries, J. Ward, S. Valle, G. Stephens, P. Moselund, K. V. D. Zanden, and B. Napier, "Mid infrared hyper-spectral imaging with bright super continuum source and fast acousto-optic tuneable filter for cytological applications," *Journal of Physics: Conference Series* **619**, 12032 (2015).

[16] T. Yuan, M. A Pleitez, F. Gasparin, V. Ntziachristos, "Wide-Field Mid-Infrared Hyperspectral Imaging by Snapshot Phase Contrast Measurement of Optothermal Excitation," *Analytical Chemistry* **93**, 15323-15330 (2021).

[17] X. Ge, Y. Zhu, D. Sun, H. Sun, Y. Li, C. V. P. Dessai, J.-X. Cheng, "Fiber laser based stimulated Raman photothermal microscopy towards a high-performance and user-friendly chemical imaging platform," *PhotoniX* **6**, 35 (2025).

[18] J. Yin, M. Zhang, Y. Tan, Z. Guo, H. He, L. Lan, J.-X. Cheng, "Video-rate mid-infrared photothermal imaging by single-pulse photothermal detection per pixel," *Sci. Adv.* **9**, eadg8814 (2023).

[19] D. Knez, B. W. Toulson, A. Chen, M. H. Ettenberg, H. Nguyen, E. O. Potma, and D. A. Fishman, "Spectral imaging at high definition and high speed in the mid-infrared," *Sci. Adv.* **8**, eade4247 (2022).

[20] S. Junaid, S. C. Kumar, M. Mathez, M. Hermes, N. Stone, N. Shepherd, M. Ebrahim-Zadeh, P. Tidemand-Lichtenberg, and C. Pedersen, "Video-rate, mid-infrared hyperspectral upconversion imaging," *Optica* **6**, 702-708 (2019).

[21] Y. Zhao, S. Kusama, Y. Furutani, W.-H. Huang, C.-W. Luo, and T. Fuji, "High-speed scanless entire bandwidth mid-infrared chemical imaging," *Nat. Commun.* **14**, 3929 (2023).

[22] S. Yang, X. Yan, H. Qin, Q. Zeng, Y. Liang, H. Arguello, X. Yuan, "Mid-Infrared Compressive Hyperspectral Imaging," *Remote sensing* **13**, 741 (2021).

[23] S. Yang, H. Qin, X. Yan, S. Yuan, Q. Zeng, "Mid-Wave Infrared Snapshot Compressive Spectral Imager with Deep Infrared Denoising Prior," *Remote Sensing* **15**, 19 (2023).