

Coherent Optical Modems for Full-Wavefield Lidar Supplemental Document

PARSA MIRDEHGHAN, University of Toronto and Vector Institute, Canada
 BRANDON BUSCAINO, Ciena Corporation, USA
 MAXX WU, University of Toronto and Vector Institute, Canada
 DOUG CHARLTON, Ciena Corporation, Canada
 MOHAMMAD E. MOUSA-PASANDI, Ciena Corporation, Canada
 KIRIAKOS N. KUTULAKOS, University of Toronto and Vector Institute, Canada
 DAVID B. LINDELL, University of Toronto and Vector Institute, Canada

S1 Coherent Modulation and Demodulation

Homodyne detection. For completeness, we provide a detailed mathematical description of the homodyne detection procedure that was introduced in Section 3 of the paper. Our derivation follows that of Ip et al. [2008]. However, for simplicity, we will ignore the effects of phase noise and amplified spontaneous emission noise in this derivation; these can be gathered into a single term consisting of complex Gaussian noise as we will note later.

Recall from the main text that the transmit and received electric fields can be written as

$$E_{\text{TX}}(t) = \sqrt{P_{\text{TX}}} e^{j\omega t} \mathbf{X}(t), \quad (\text{S1})$$

$$\begin{aligned} E_{\text{RX}}(t) &= \mathbf{R} E_{\text{TX}}(t - \tau) e^{j\nu t} \\ &= \sqrt{P_{\text{TX}}} \mathbf{R} \mathbf{X}(t - \tau) e^{j\nu t} e^{j\omega t}, \end{aligned} \quad (\text{S2})$$

where P_{TX} is the transmit optical power. Then, write the electric field corresponding to the local oscillator as

$$E_{\text{LO}}(t)[p] = \sqrt{P_{\text{LO}}} e^{j\omega t}, \quad p \in \{1, 2\}, \quad (\text{S3})$$

where p indexes the polarization channel and P_{LO} is the transmit power of the local oscillator.

As shown in Figure S1, the received and local oscillator fields are combined and detected using two pairs of balanced photodiodes—one pair for each polarization channel. The photocurrent $I[p]$ at the output of each balanced photodiode is given as

$$I[p] = |E_{\text{RX}}(t)[p] + E_{\text{LO}}(t)[p]|^2 - |E_{\text{RX}}(t)[p] - E_{\text{LO}}(t)[p]|^2, \quad (\text{S4})$$

where the sign change in the second term comes from a 90° phase shifter and an additional 90° phase shift induced by the fiber coupler (see Figure S1).

Expanding the first term (and dropping dependencies on t for convenience) yields

$$\begin{aligned} &|E_{\text{RX}}[p] + E_{\text{LO}}[p]|^2 \\ &= (\overline{E_{\text{RX}}[p] + E_{\text{LO}}[p]})(E_{\text{RX}}[p] + E_{\text{LO}}[p]) \\ &= P_{\text{TX}} |(\mathbf{R}\mathbf{X})[p]|^2 + P_{\text{LO}} + (\overline{E_{\text{RX}}E_{\text{LO}}})[p] + (E_{\text{RX}}\overline{E_{\text{LO}}})[p]. \end{aligned} \quad (\text{S5})$$

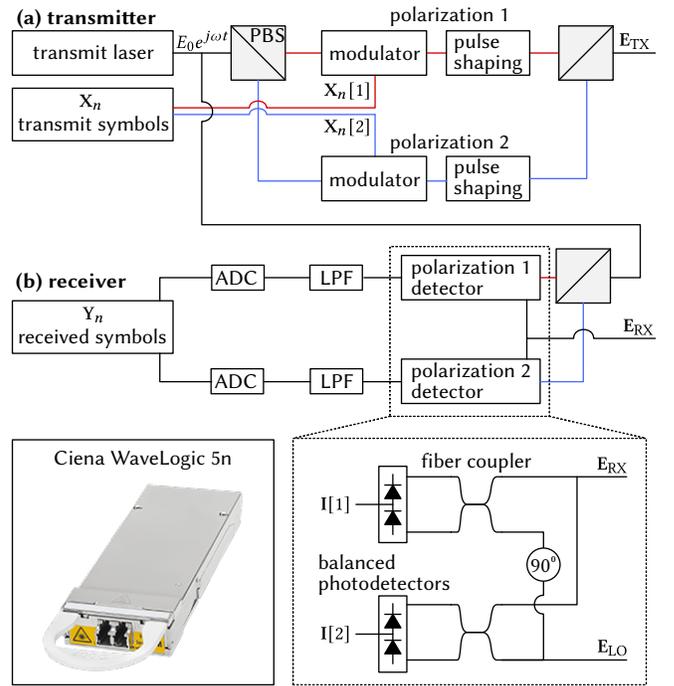


Fig. S1. Illustration of the transmit and receive path of a coherent optical modem. (a) The transmit signal comprises two polarization channels that are modulated using the transmit symbol sequence X_n , pulse-shaped, and combined using a polarizing beam splitter (PBS). (b) The received signal is mixed with the local oscillator to perform downconversion and homodyne detection (inset, bottom right). Balanced photodetectors convert the intensity in each polarization channel to a photocurrent $I[p]$ that is low-pass filtered (LPF) and sampled using an analog-to-digital converter (ADC) to capture the received symbol sequence Y_n . A photo of a coherent optical modem is shown in the bottom left.

Expanding the second term in Equation S4, we similarly obtain

$$\begin{aligned} &|E_{\text{RX}}[p] - E_{\text{LO}}[p]|^2 \\ &= P_{\text{TX}} |(\mathbf{R}\mathbf{X})[p]|^2 + P_{\text{LO}} - (\overline{E_{\text{RX}}E_{\text{LO}}})[p] - (E_{\text{RX}}\overline{E_{\text{LO}}})[p]. \end{aligned} \quad (\text{S6})$$

Subtracting Equation S6 from Equation S5 gives

$$\begin{aligned} I[p] &= 2(\overline{\mathbf{E}_{\text{RX}}\mathbf{E}_{\text{LO}}})[p] + 2(\mathbf{E}_{\text{RX}}\overline{\mathbf{E}_{\text{LO}}})[p] \\ &= 2\sqrt{P_{\text{TX}}}\sqrt{P_{\text{LO}}}|(\mathbf{R}\mathbf{X})[p]|\left(e^{-j(vt+\angle(\mathbf{R}\mathbf{X})[p])} + e^{j(vt+\angle(\mathbf{R}\mathbf{X})[p])}\right) \\ &= \underbrace{4\sqrt{P_{\text{TX}}}\sqrt{P_{\text{LO}}}|(\mathbf{R}\mathbf{X})[p]|}_{A[p]}\cos(\underbrace{vt + \angle(\mathbf{R}\mathbf{X})[p]}_{\phi[p]}). \end{aligned} \quad (\text{S7})$$

So the balanced photodetectors remove all steady-state signals. Any terms due to ambient light would also be canceled out in the balanced photodetection. The final photodetector current is a cosine function whose amplitude is the product of the transmit and receive amplitudes, the Jones matrix, and the transmitted symbol amplitudes. The frequency depends on the Doppler shift, and the phase of the signal depends on the phase of the Jones matrix entries and the transmitted symbols.

In practice, the coherent modem captures complex-valued samples of the photocurrent. That is, the photocurrent signal is passed through a splitter, and one signal copy is sampled directly while the other copy is delayed with a 90 degree phase shift and then sampled. The resulting signal is

$$\underbrace{A[p]\cos(vt + \phi[p])}_{\text{in-phase}} + \underbrace{A[p]\sin(vt + \phi[p])}_{\text{quadrature}}. \quad (\text{S8})$$

These two sampled signals are commonly called the in-phase and quadrature signal components, where the quadrature signal corresponds to the phase-delayed copy.

Finally, treating the in-phase and quadrature signals as the real and imaginary components of a complex-valued signal, respectively, yields

$$\begin{aligned} &A[p]\cos(vt + \phi[p]) + jA[p]\sin(vt + \phi[p]) \\ &= A[p]e^{j(vt+\phi[p])} \\ &\propto \mathbf{E}_{\text{RX}}[p]e^{-j\omega t}. \end{aligned} \quad (\text{S9})$$

such that the homodyne detection procedure recovers a complex-valued signal that is proportional to the received electric field. While we ignore noise in this derivation for simplicity, it is typically modeled using a complex Gaussian distribution. The dominating sources of noise are local oscillator shot noise and amplified spontaneous emission noise due to the Erbium-doped fiber amplifiers used to amplify the received laser light [Ip et al. 2008].

S2 Supplemental Implementation Details

Optimization. We implement the optimization using a two-stage procedure. First, we note that in the absence of total variation regularization the objective function (Equation 13) can be minimized in a per-pixel fashion. For computational expediency, we conduct a first stage of optimization where $\mathbf{R}_{\Delta, \nu}$ are estimated for each pixel in parallel using only the sparsity regularizer. We also assume the maximum plausible distance from the system to be 4 meters, and only optimize for the Jones matrices associated with the feasible time delays. We find that 50 iterations of optimization using Adam [Kingma and Ba 2015] with $\lambda_{\text{sparse}} = 10^{-1}$ for static scenes and $\lambda_{\text{sparse}} = 3 \times 10^{-1}$ for dynamic scenes is sufficient for the estimated depth to converge. To avoid unnecessary computation during

the optimization, we assume that $\mathbf{R}_{\Delta, \nu} = 0$ for all $\nu \neq 0$ for scenes that are known to be static (i.e., no Doppler shift). In this case we optimize only the set of Jones matrices for which $\nu = 0$. For static pixels, this optimization requires a few seconds per pixel using an NVIDIA A40 GPU. With our unoptimized implementation, processing each pixel with a Doppler shift requires roughly one minute on the same hardware.

In the second stage of optimization we add the total variation penalty; this procedure requires processing the entire image at once due to the dependencies between pixels. However, given the long sequence lengths of $\mathbf{R}_{\Delta, \nu}$ (typically several thousand samples along the temporal dimension), $E_{\text{TX}} (\approx 2^{16}$ symbols), plus the additional dimensions associated with the number of pixels and the entries of the Jones matrix, it is challenging to process the entire captured dataset at every iteration using full-batch gradient methods. Instead, we stochastically sample a number of pixels and their neighbors to calculate each term of Equation 13, including the data term, the sparsity term, and the total variation penalty (we use a batch size of 1024 pixels at each iteration). We find that this stage of the optimization converges within 500 iterations, for a total of 550 iterations of optimization including the first stage. The second stage of optimization takes roughly 30 minutes to complete using an Nvidia A40 GPU.

Finally, we note that we do not apply the total variation penalty to dynamic scenes when estimating the Doppler shifts for the Jones matrices. While we find the total variation penalty to be effective for the static scenes (e.g., Figure 9), the additional velocity dimension increases memory requirements, making it challenging to apply total variation to the large amount of data captured by the optical modem across all possible frequency shifts. Finding efficient ways to work with the large quantities of data captured by an optical modem is an interesting direction for future work.

Implementation of other modulation schemes. We compare FWL to using phase-only modulation with two polarization channels, amplitude-only modulation with two polarization channels, and phase and amplitude modulation with one polarization channel. The transmit power is kept the same as for FWL in all cases (including for the single polarization modulation scheme).

To implement the phase-only modulation scheme, we generate a symbol sequence with constant amplitude and uniformly distributed phases. Since our intent is to emulate phase modulation-based lidars that are not sensitive to amplitude, we discard the amplitude information by normalizing the symbols at the receiver prior to estimating depth or velocity. For amplitude-only modulation, we transmit symbols with constant phase and normally distributed amplitudes. The reconstruction is performed by projecting the received symbols onto a complex-valued unit vector with same phase used for modulation. This procedure removes the phase information from the receiver. Finally, for single-polarization phase and amplitude modulation, we simply discard one of the polarization channels at the receiver.

Single-photon lidar system. To compare FWL with single-photon lidar (Figure 2), we built a prototype single-photon lidar system, shown in Figure S2. The prototype comprises a single-pixel single-photon avalanche diode (SPAD; MPD Fast-Gated Module), a beam

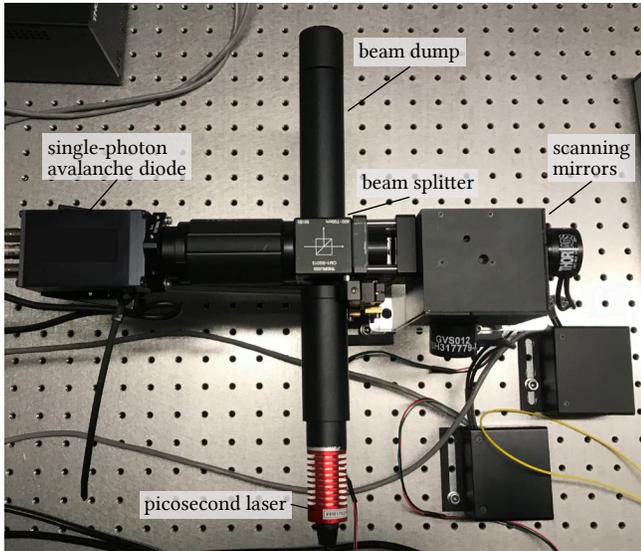


Fig. S2. Prototype single-photon lidar system. A single-pixel, single-photon avalanche diode shares an optical path through a beamsplitter with a picosecond laser. A pair of scanning mirrors is used to raster scan the scene.

splitter to separate the outgoing laser beam and incoming reflected light, a picosecond pulsed laser operating at 670 nm (Alphas Pi-copower), a pair of scanning mirrors (Thorlabs GVS012), and a time-correlated single photon counter (TCSPS; PicoQuant PicoHarp 300). We set the laser pulse repetition rate to 10 MHz and configure the power to achieve approximately 500,000 photon counts per second, which is roughly the threshold at which non-linear pileup effects are still negligible [Rapp et al. 2021].

S3 Supplemental Results

Supplemental ablation studies. We provide additional results showing the performance of FWL with both total variation and sparsity regularization, without total variation regularization, and without any regularization (Figure S3). We find that using both regularizers produces the best results for this static scene. We also compare to the performance of generalized matched filtering, which correlates the known transmit sequence at each polarization channel with each of the received polarization channels (i.e., exploits cross-channel polarization information). We find that this approach recovers depth estimates with fewer outliers than a naive form of matched filtering, which only correlates the corresponding transmit and receive polarization channels (also shown in Figure S3). We observe similar trends in Figure S4, which shows the same comparison for a range of exposure times. As the exposure time decreases, we find that FWL using both sparsity and total variation regularization produces depth maps with the fewest outliers compared to matched filtering or FWL without sparsity or total variation regularization.

Quantitative evaluation of estimated radial velocity. To evaluate the estimated radial velocities shown in Figure 1 and Figure 9 of the main paper, we measure the rotational speed of the fan using a high speed camera. Then, we fit a plane to the scene, which contains a

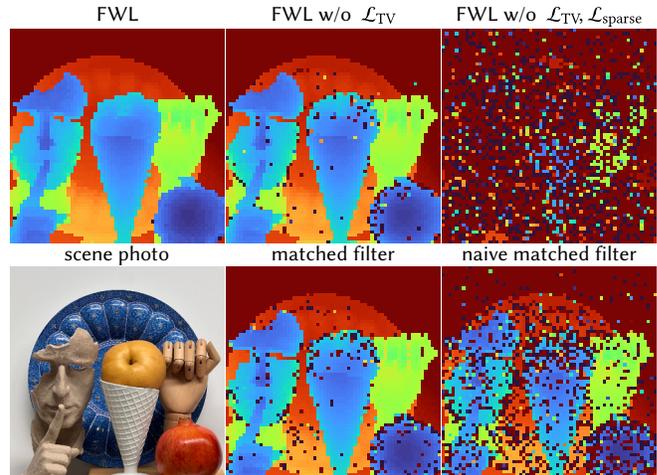


Fig. S3. Performance of FWL regularization and matched filtering. FWL using both total variation and sparsity regularizers (top left) performs best for this scene containing a few static objects. The method without total variation (top middle) is more sensitive to speckle noise, and removing both regularizers results in poor performance. The generalized matched filter (bottom middle) performs better than “naive” matched filtering (bottom right), which uses cross correlations between the two corresponding polarization channels and fails to exploit cross-polarization information.

spinning disk that we retrofit to a fan motor. Using the plane fit, we compute the surface normal corresponding to each measured pixel, and we estimate the per-pixel radial velocities. We find that this approach for estimating velocity agrees with our estimates using the Doppler shift to within a meter per second in terms of mean absolute error (MAE) as shown in Figure S5. Measurements from each approach are compared using the same spanning disk at different orientations and for different motor speed settings.

Illustration of Figure 1 setup. We provide a labeled image depicting the capture setup used for Figure 1. In particular, we note that the scene was captured near a window and was illuminated by strong ambient light from the sun.

References

- Ezra Ip, Alan Pak Tao Lau, Daniel JF Barros, and Joseph M Kahn. 2008. Coherent detection in optical fiber systems. *Opt. Express* 16, 2 (2008), 753–791.
- Diederik Kingma and Jimmy Ba. 2015. Adam: A Method for Stochastic Optimization. In *Proc. ICLR*.
- Joshua Rapp, Yanting Ma, Robin MA Dawson, and Vivek K Goyal. 2021. High-flux single-photon lidar. *Optica* 8, 1 (2021), 30–39.

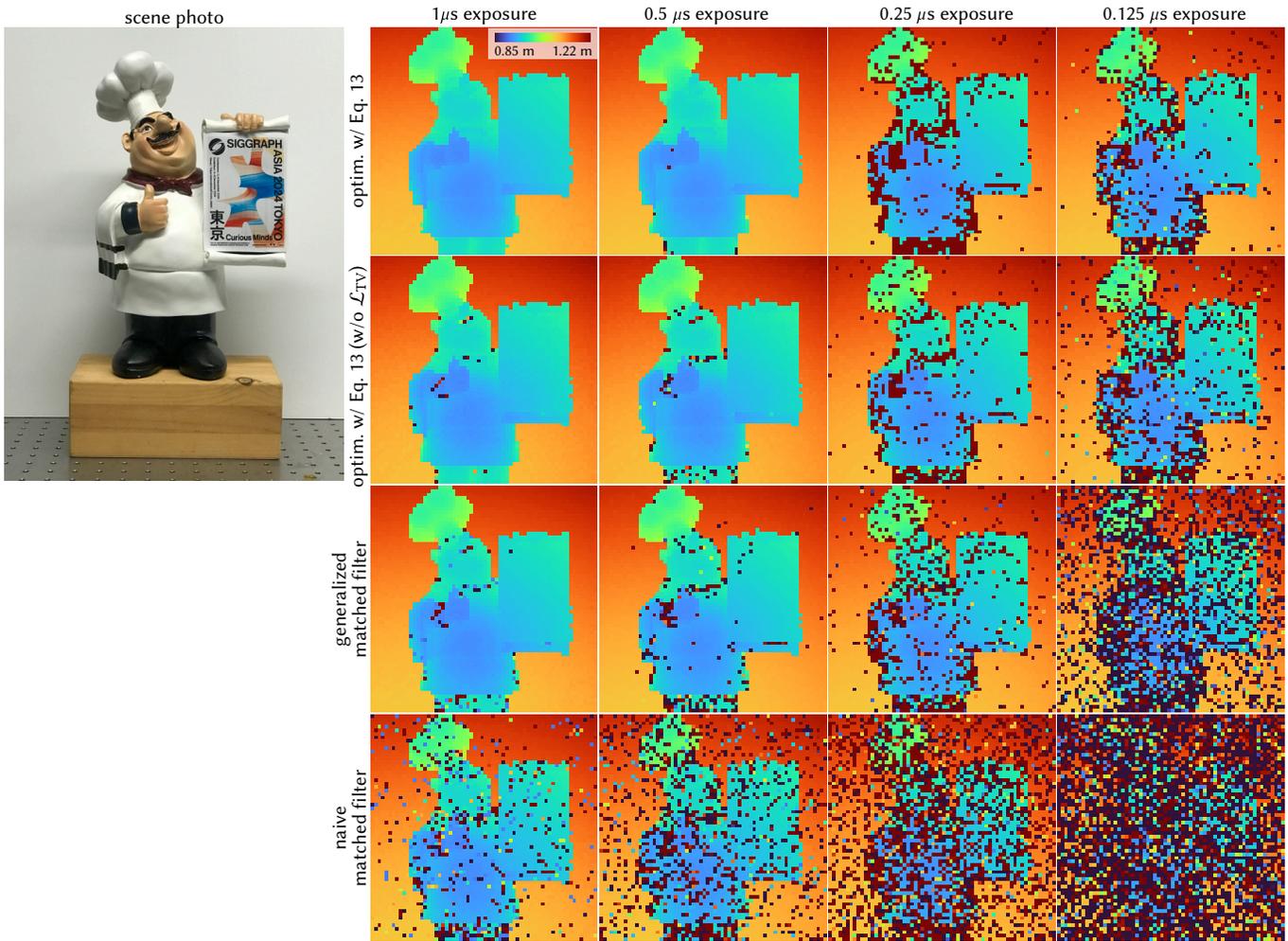


Fig. S4. Additional comparisons of FWL and matched filtering vs. exposure time. We compare reconstruction using both sparsity and total variation regularization (row 1) to the optimization without total variation regularization (row 2). We also compare to generalized matched filtering (row 3), which exploits cross polarization information, and naive matched filtering (row 4), which only correlates the two corresponding polarization channels. FWL using both regularizers produces depth maps with the fewest outliers as the exposure time decreases.

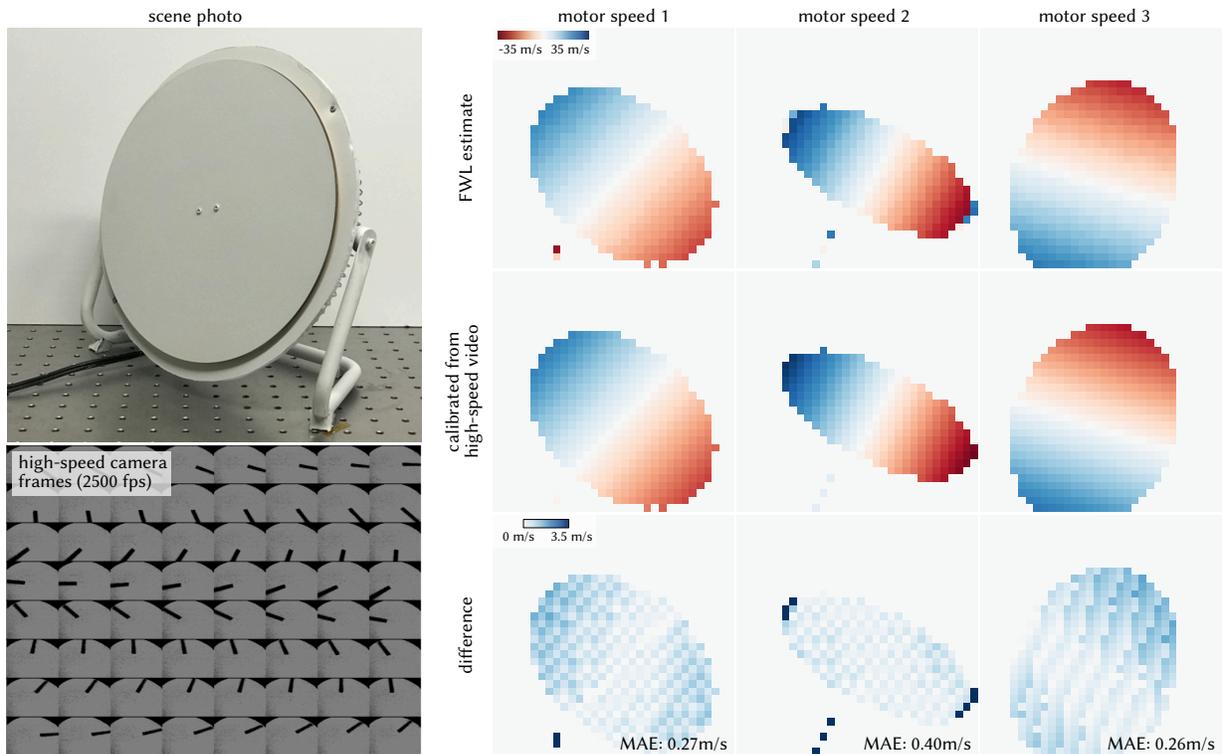


Fig. S5. Assessment of the radial velocity predictions from the coherent optical modem for a spinning disk scene. We use a high-speed camera to capture frames of the spinning disk after adding a marker to track its position. Then, we fit a plane to the depth estimates from the coherent optical modem and use the rotational speed from the high-speed camera to estimate the radial velocity of the disk. We show the radial velocities computed in this fashion and using the Doppler shift from the optical modem for three different fan orientations and motor speeds. We find that the estimates from both methods agree to within a meter per second in terms of mean absolute error (MAE; right, bottom row). We use a transmit power of 10 mW for this experiment.

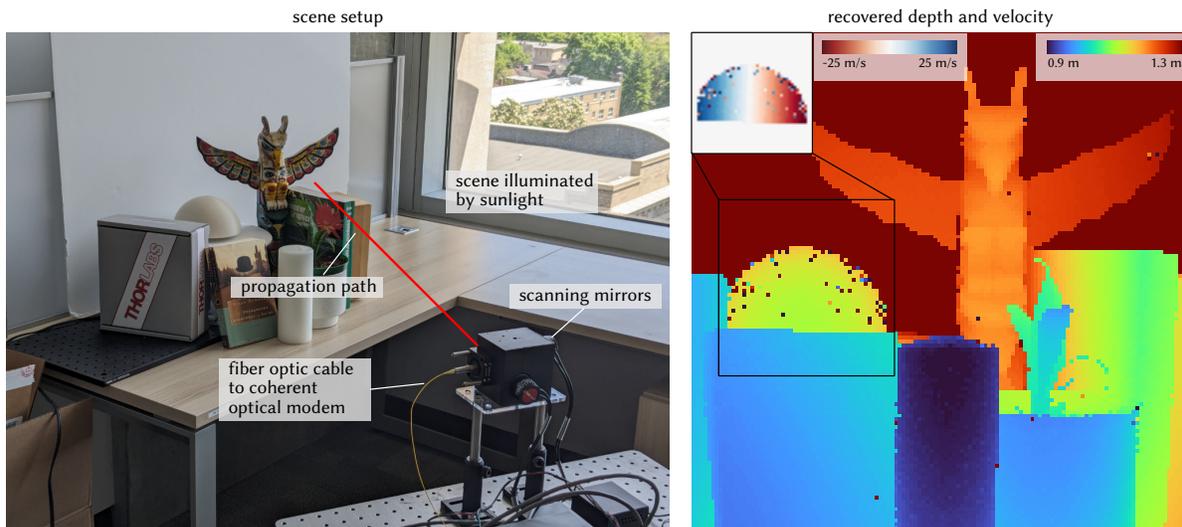


Fig. S6. Capture setup of Figure 1. We show a labeled photo (left) depicting the capture setup for Figure 1, including illumination by sunlight through a window. The fiber optic cable to the modem, scanning mirrors, and propagation path are also labeled. For convenience, the depth and velocity reconstructions are reproduced from the main paper (right).