Supplementary Materials for

Large-array sub-millimeter precision coherent flash three-dimensional imaging

The file includes:

Supplementary Note 1: Mathematical model

An optical signal is generated from a laser diode and sent to a high-speed Mach-Zehnder modulator (MZM). The electrical field of the optical signal can be expressed as:

$$E_0(t) = Ae^{i\omega t} \tag{1}$$

where A is the amplitude and ω is the angular frequency of the optical signal.

The input microwave signal can be expressed as:

$$V(t) = V_m cos\left(2\pi f_m t\right) \tag{2}$$

where V_m and f_m are the amplitude and frequency of the microwave signal.

In the MZM modulator, the microwave signal is loaded on the optical carrier in one arm and a DC voltage signal is applied to tune the phase. Therefore, the phase difference between the optical signals in the two arms is:

$$\Delta\varphi(t) = \frac{\pi V_m}{V_\pi} cos\left(2\pi f_m t\right) + \frac{\pi V_{bias}}{V_\pi} \tag{3}$$

where V_π and V_{bias} are the half-wave voltage and bias voltage of the MZM.

At the output of the MZM, the optical signal can be written as:

$$E_{MZM}(t) = \frac{\sqrt{2}}{2} A\left(e^{i\Delta\varphi(t)} + 1\right)e^{i\omega t} \tag{4}$$

One portion of this signal is employed to flood illuminate the scene, while the remaining portion serves as a local reference signal. When the transmitted optical signal meets the target, there would be an echo signal. The electrical field of the local and echo signal can be expressed:

$$E_l(t) = \frac{\sqrt{2}}{2} A_l \left[e^{i\Delta\varphi(t)} + 1\right] e^{i\omega t} \tag{5}$$

$$E_e(t) = \frac{\sqrt{2}}{2} A_e \left[e^{i\Delta\varphi(t-\tau)} + 1\right] e^{i\omega(t-\tau)} \tag{6}$$

where $\tau = 2d / c$ is the time delay of the echo signal relative to the local signal, d is the depth of the target, c is the speed of light, A_l and A_e are the amplitude of local signal and echo signal.

When the local signal and the echo signal come into the coherent image sensor, at the output of each pixel photodetector, the generated voltage can be written as[S1]:

$$U = \frac{2\pi\eta}{h\omega} S_V G T_e \left\langle \left[E_l^*(t) + E_e^*(t)\right]\left[E_l(t) + E_e(t)\right] \right\rangle \tag{7}$$

where η is the quantum efficiency, h is the Planck constant, S_V is the sensitivity of the sense node, G is the gain of CCD pixels and T_e is the exposure time of each frame.

By substituting Eqn. (5) and (6) into Eqn. (7) and using the Bessel expansion, considering small-signal modulation and ignoring high frequency components, the generated voltage can be approximately rewritten as:

$$U = \frac{\pi\eta}{h\omega} S_V G T_e \left[A_l^2 + A_e^2 + \left(A_e^2 + A_l^2 \right) J_0(\beta) \cos\left(\frac{\pi V_{bias}}{V_\pi} \right) \right] + \frac{4\pi\eta}{h\omega} S_V G T_e A_l A_e J_1^2(\beta) \cos(2\pi f_m \tau) \cos(\omega\tau) \tag{8}$$

where J denotes Bessel function of the first kind and β is the modulation index. From Eqn. (8), the generated voltage includes a DC term and a multiplication component. Thanks to the coherent detection, at the output of each pixel of four CCDs, the generated voltage of each pixel at four CCDs can be written as:

$$\begin{aligned}
U_1 &= U_{DC} + \frac{4\pi\eta}{h\omega} S_V G T_e A_l A_e J_1^2(\beta) cos\left(\frac{4\pi f_m d}{c} \right) cos\left(\frac{2\omega d}{c} - \pi \right) \\
U_2 &= U_{DC} + \frac{4\pi\eta}{h\omega} S_V G T_e A_l A_e J_1^2(\beta) cos\left(\frac{4\pi f_m d}{c} \right) cos\left(\frac{2\omega d}{c} - \frac{\pi}{2} \right) \\
U_3 &= U_{DC} + \frac{4\pi\eta}{h\omega} S_V G T_e A_l A_e J_1^2(\beta) cos\left(\frac{4\pi f_m d}{c} \right) cos\left(\frac{2\omega d}{c} + \frac{\pi}{2} \right) \\
U_4 &= U_{DC} + \frac{4\pi\eta}{h\omega} S_V G T_e A_l A_e J_1^2(\beta) cos\left(\frac{4\pi f_m d}{c} \right) cos\left(\frac{2\omega d}{c} \right)
\end{aligned} \tag{9}$$

where $U_{DC} = \frac{\pi\eta}{h\omega} S_V G T_e \left[A_l^2 + A_e^2 + \left(A_e^2 + A_l^2 \right) J_0(\beta) \cos\left(\frac{\pi V_{bias}}{V_\pi} \right) \right]$ is the DC term. The DC term is a constant for each CCD. By combining the output of the four CCDs, the multiplication component can be reached:

$$\begin{aligned}
U_{sig} &= \sqrt{(U_1 - U_4)^2 + (U_2 - U_3)^2} \\
&= \frac{8\pi\eta}{h\omega} S_V G T_e A_l A_e J_1^2(\beta) cos\left(\frac{4\pi f_m d}{c} \right)
\end{aligned} \tag{10}$$

To realize the depth measurement, a stepped-frequency microwave signal is applied to the MZM. In the frequency domain, the instantaneous frequency of the microwave signal can be written as[S2]:

$$f_m = f_0 + k\Delta f,\ k = 0, 1, \cdots, \mathrm{N} - 1 \tag{11}$$

where f_0 is the start frequency of microwave signal, Δf is the frequency step size, and N is the step points. Each frequency has a temporal duration of T_{step}.

Substituting Eqn. (11) into Eqn. (10), the equation can be rewritten as:

$$U_{sig}(k)=\frac{8\pi\eta}{h\omega}S_V G T_e A_l A_e J_1^2(\beta)\cos\left[\frac{4\pi d}{c}\left(f_0+k\Delta f\right)\right],\ k=0,1,\cdots,\mathrm{N}-1 \tag{12}$$

After performing discrete Fourier transform (DFT) and taking the magnitude, Eqn. (12) can be rewritten in the frequency domain:

$$\left|\mathcal{U}_{sig}(n)\right|=\frac{4\pi\eta}{h\omega}S_V G T_e A_l A_e J_1^2(\beta)\left|\frac{\sin\left[\pi\left(n-\frac{2N\Delta fd}{c}\right)\right]}{\sin\left[\frac{\pi}{N}\left(n-\frac{2N\Delta fd}{c}\right)\right]}\right| \tag{13}$$

Based on Eqn. (13), the depth d can be retrieved by extracting the fundamental frequency:

$$d=\frac{c}{2N\Delta f}n_{funda} \tag{14}$$

where n_{funda} is the fundamental frequency.

The range resolution can be written as:

$$\Delta d=\frac{c}{2N\Delta f}=\frac{c}{2B} \tag{15}$$

where B is the bandwidth of the stepped-frequency signal. As can be seen, the broader the bandwidth, the higher the resolution.

The unambiguous range can be expressed as:

$$d_{NAR}=\frac{c}{4\Delta f} \tag{16}$$

As can be seen, the smaller the frequency step size, the larger the unambiguous range.

Supplementary Note 2: 2D image reconstruction

In the experiment, the local optical signal has a power of 14.12 μW. After passing through the free-space optical components, the power of the local signal is reduced to be 0.84 μW. Considering 320 × 256 pixels of each CCD, the power that each pixel receives is calculated to be 2.54 pW.

The probe optical signal has a power of 12 dBm, approximately 15.85 mW. After the target backscattering, the received echo power can be estimated via the LiDAR equation:

$$P_e=P_p R\frac{\pi D^2}{4\Omega d^2}\eta_{atm}\eta_{sys} \tag{17}$$

where P_p is the optical power, R is the reflectivity of the target, D is the aperture of the receiving lens, Ω is the solid angle of scattering, η_{atm} is the transmission efficiency, and η_{sys} is the system efficiency.

Assuming the best situation with a full reflectivity and a full efficiency, the received echo power is estimated to be 334.88 pW. Thus, the power that each pixel receives is calculated to be 1.02 fW, which is three orders of magnitude lower than that of the local signal.

Based on Eqn. (8), the voltage signal at the output of each pixel includes a DC component and a multiplication component. When a stepped-frequency microwave signal is used to modulate the optical carrier, the multiplication component has a cosine relationship with the microwave signal frequency. By performing the average operation, the multiplication component can be ignored. Therefore, considering $A_l \gg A_e$, Eqn. (8) can be rewritten as:

$$\begin{aligned}\langle U\rangle &\approx \frac{\pi\eta}{h\omega} S_V G T_e \left[A_l^2 + A_e^2 + \left(A_e^2 + A_l^2\right) J_0(\beta)\cos\left(\frac{\pi V_{bias}}{V_\pi}\right)\right] \\ &\approx \frac{\pi\eta}{h\omega} S_V G T_e \left[1 + J_0(\beta)\cos\left(\frac{\pi V_{bias}}{V_\pi}\right)\right] A_l^2 \\ &= \frac{\pi\eta}{h\omega} S_V G T_e \left[1 + J_0(\beta)\cos\left(\frac{\pi V_{bias}}{V_\pi}\right)\right] P_l\end{aligned} \tag{18}$$

After averaging, the result is proportional to the local optical signal power. From Eqn. (13), at the fundamental frequency, the multiplication component is proportional to $\left|\mathcal{U}_{sig}\left(n_{funda}\right)\right|$:

$$P_l P_e = \left(A_l A_e\right)^2 \propto \mathcal{U}_{sig}^2\left(n_{funda}\right) \tag{19}$$

Therefore, by combining Eqn. (18) and (19), the echo signal power can be arrived:

$$P_e \propto \frac{\mathcal{U}_{sig}^2}{\langle U\rangle} \tag{20}$$

Using Eqn. (20), the 2D image reconstruction is present in Fig. 3f. For comparison, Fig. S3 gives the averaged result of the original image captured by each CCD in total 3000 frames. As can be seen, due to the ultra-small echo power, there is no visible 2D target image at the output of each CCD. In addition, in the figure, some speckles can be found which is caused by the random interference.

Therefore, the coherent detection process enhances the ability to detect extremely small echo signals by amplifying the multiplication component through the local signal. This amplification significantly improves the sensitivity of the system, making it possible to capture and analyze very weak signals. Consequently, this advancement allows for the creation of high-resolution 3D images over extended distances.

Supplementary Note 3: Theoretical range precision analysis

To assess the range precision of the proposed LiDAR system, we use the Cramér-Rao lower bound (CRLB) as a benchmark[S3]. The CRLB provides a theoretical lower limit on the precision with which the range can be estimated. Typically, the CRLB sets the limit for frequency precision when dealing with a time-limited sine wave in the presence of white Gaussian noise. The formula for this lower bound in frequency precision is:

$$\mathrm{var}(\hat{f}_{sig}) \geq \frac{3 f_{CCD}^2}{\pi^2 N(N-1)(2N-1)} \cdot \frac{1}{SNR} \tag{21}$$

where N is the step points and SNR is the signal to noise ratio of the output voltage signal of the coherent image sensor. For $N \gg 1$ and $f_{CCD} = N/T$, Eqn. (21) can be rewritten as:

$$\mathrm{var}(\hat{f}_{sig}) \geq \frac{3}{2\pi^2 T^2 N} \cdot \frac{1}{SNR} \tag{22}$$

Therefore, the frequency precision of the signal affected by noise can be calculated using the standard deviation:

$$\delta f_{sig} \geq \frac{1}{\pi T}\sqrt{\frac{3}{2N} \cdot \frac{1}{SNR}} \tag{23}$$

Combining Eqn. (23) and Eqn. (14), the range precision δd can be written as:

$$\delta d \geq \frac{\mathrm{c}}{2\pi \mathrm{B}}\sqrt{\frac{3}{2N} \cdot \frac{1}{SNR}} \tag{24}$$

From Eqn. (24), the range precision depends on the bandwidth, the number of step points, and the SNR of the output voltage signal. Generally, the broader the bandwidth, the higher the range precision.

The SNR of the output voltage signal is a key factor to determine the range precision. From Eqn. (8), the accumulated charge at the output of each pixel can be written as:

$$\begin{aligned} Q = \frac{U}{S_V G} &= \frac{\pi\eta}{h\omega} T_e \left[A_l^2 + A_e^2 + \left(A_e^2 + A_l^2\right) J_0(\beta) \cos\left(\frac{\pi V_{bias}}{V_\pi}\right) \right] \\ &+ \frac{4\pi\eta}{h\omega} T_e A_l A_e J_1^2(\beta) \cos\left(2\pi f_m \tau\right) \cos\left(\omega\tau\right) \end{aligned} \tag{25}$$

The accumulated charge includes a DC component and a multiplication component. Since $A_l \gg A_e$, the DC component can be written as:

$$Q_{DC} \approx \frac{\pi\eta}{h\omega} T_e \left[1 + J_0(\beta) \cos\left(\frac{\pi V_{bias}}{V_\pi}\right) \right] A_l^2 \tag{26}$$

Meanwhile, the multiplication component can be written as:

$$Q_{sig} = \frac{4\pi\eta}{h\omega} T_e A_l A_e J_1^2(\beta) \cos\left(2\pi f_m \tau\right) \cos\left(\omega\tau\right) \tag{27}$$

The average charge of the multiplication component is:

$$\begin{aligned} \left\langle Q_{sig} \right\rangle &= \frac{4\pi\eta}{h\omega} T_e A_l A_e J_1^2(\beta) \cos\left(\omega\tau\right) \left\langle \cos\left(2\pi f_m \tau\right) \right\rangle \\ &= \frac{2\sqrt{2}\pi\eta}{h\omega} T_e A_l A_e J_1^2(\beta) \cos\left(\omega\tau\right) \end{aligned} \tag{28}$$

Potential sources of noise in the system include inherent shot noise, dark current noise, and readout noise generated by the detector[S4]. The shot noise, which stems from the quantum nature of light, results from statistical variations in the number of photons emitted by the object. This type of noise is intrinsic and unavoidable in imaging systems. The shot noise Q_{n-shot} can be described using Poisson statistics[S5]:

$$Q_{n-shot} = \sqrt{Q_{DC}} \tag{29}$$

Dark current arises from thermally generated electrons in the CCD's silicon substrate. Like shot noise, the dark current noise Q_{n-dark} follows a Poisson distribution and is proportional to the square root of the number of thermal electrons generated during the exposure time.

$$Q_{n-dark} = \sqrt{\frac{i_{dark}}{e} T_e} \tag{30}$$

Therefore, the *SNR* for a single pixel can be written as:

$$SNR = \frac{\left\langle Q_{sig} \right\rangle}{\sqrt{Q_{n-shot}^2 + Q_{n-dark}^2 + Q_{n-readout}^2}} \tag{31}$$

where $Q_{n-readout}$ represents the readout noise. This type of noise is introduced by the electronic circuitry during the charge transfer process. It is primarily dependent on the frame rate but remains constant across different integration times for each frame. The SNR can be determined as follows:

$$SNR = \frac{\frac{2\sqrt{2}\pi\eta}{h\omega} T_e A_l A_e J_1^2(\beta)\cos(\omega\tau)}{\sqrt{\frac{\pi\eta}{h\omega} T_e \left[1 + J_0(\beta)\cos\left(\frac{\pi V_{bias}}{V_\pi}\right)\right] A_l^2 + \frac{i_{dark}}{e} T_e + Q_{n-readout}^2}} \tag{32}$$

Table S1 summarizes the parameters in the theoretical range precision analysis.

Supplementary Note 4: Range precision measurement

Several tests are conducted to measure range precision as the target distance varies from 15 m to 30 m, with the probe optical power held constant at 15.85 mW. Figures S4a-S4d show the 3D images of the planar target at distances of 15 m, 20 m, 25 m, and 30 m, while Figures S4E-S4H present the histograms corresponding to each 3D image. The range precision for each distance is calculated by determining the standard deviation from a Gaussian curve fitted to each histogram, yielding values of 0.42 mm, 0.44 mm, 0.45 mm, and 0.47 mm.

Supplementary Note 5: Interactive 3D reconstruction of a bust sculpture

Positioning the bust sculpture at 30.3 m away, the same 3D imaging process is performed after each rotation by 45°, and eight 3D images are collected. Figure S5 shows the eight 3D images of the bust sculpture at the different orientations, which exhibits the proper scale and perspective of the bust sculpture.

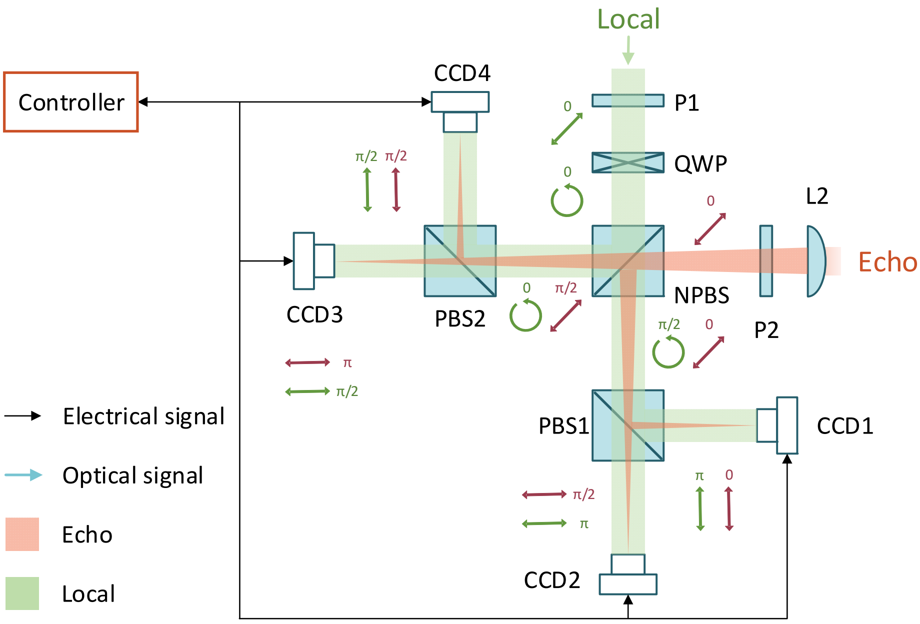

Fig. S1. Coherent image sensor. L: lens. P: polarizer. QWP: quarter-wave plate. NPBS: non-polarized beam splitter. PBS: polarized beam splitter. CCD: charge-coupled device.

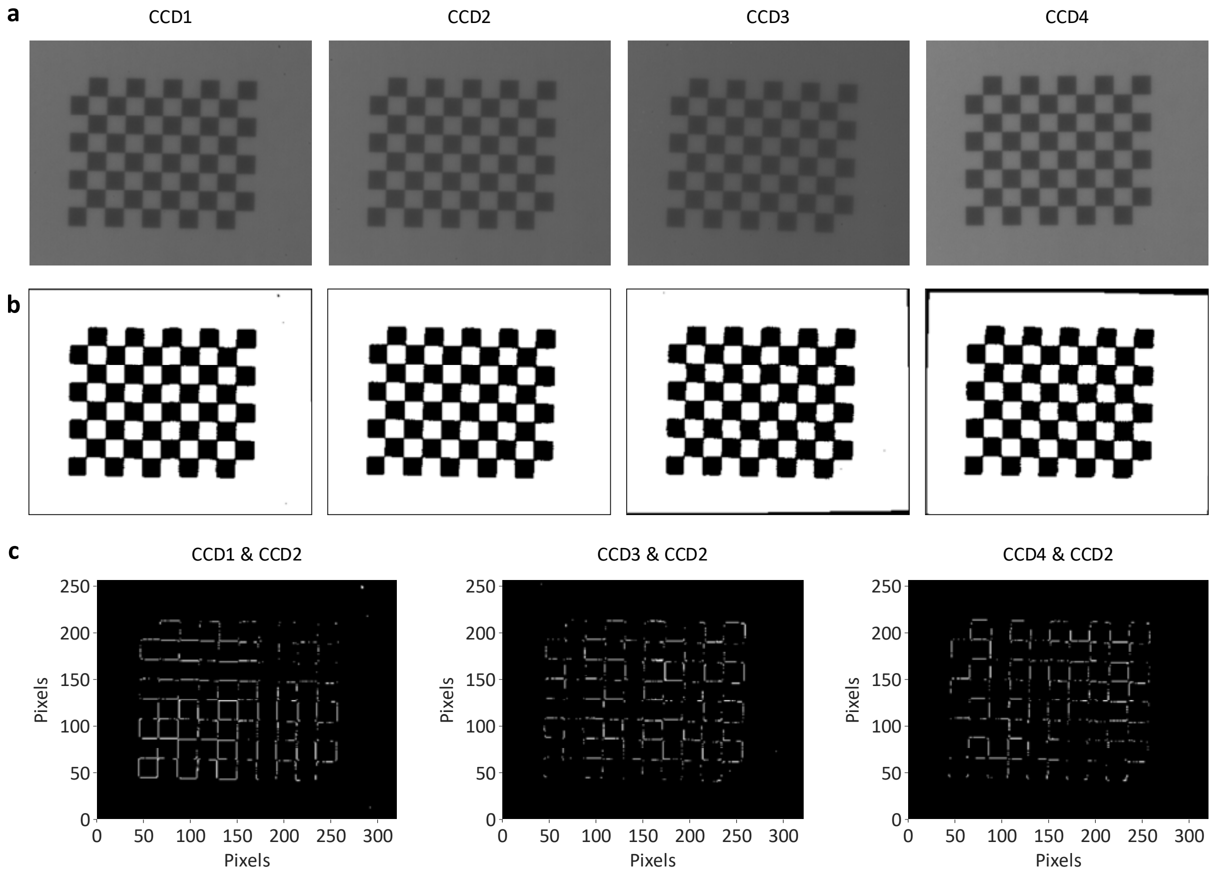

Fig. S2. Coherent image sensor calibration. a, Captured chessboard images by each CCD. **b,** Calculated chessboard images after binarization and calibration. **c,** XOR results.

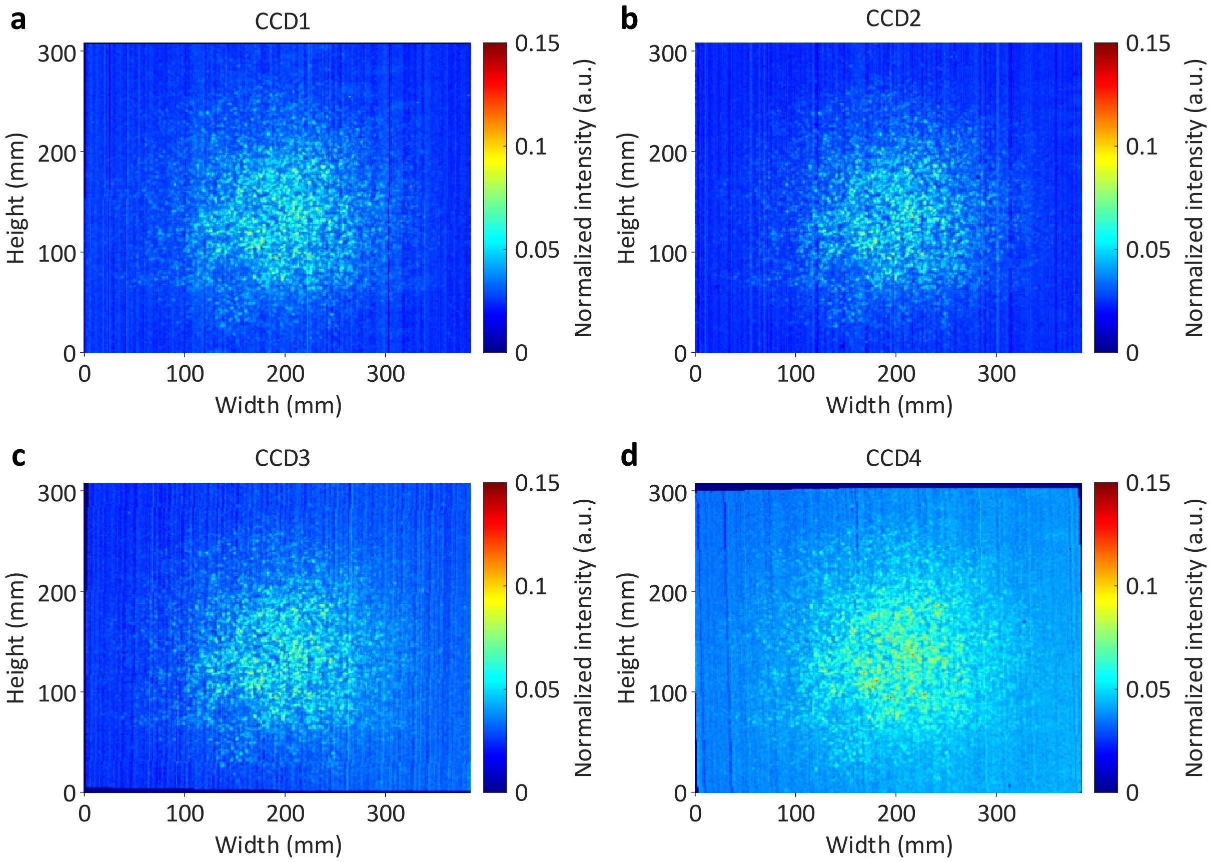

Fig. S3. Averaged results of the original images in total 3000 frames. **a**, CCD1. **b**, CCD2. **c**, CCD3. **d**, CCD4.

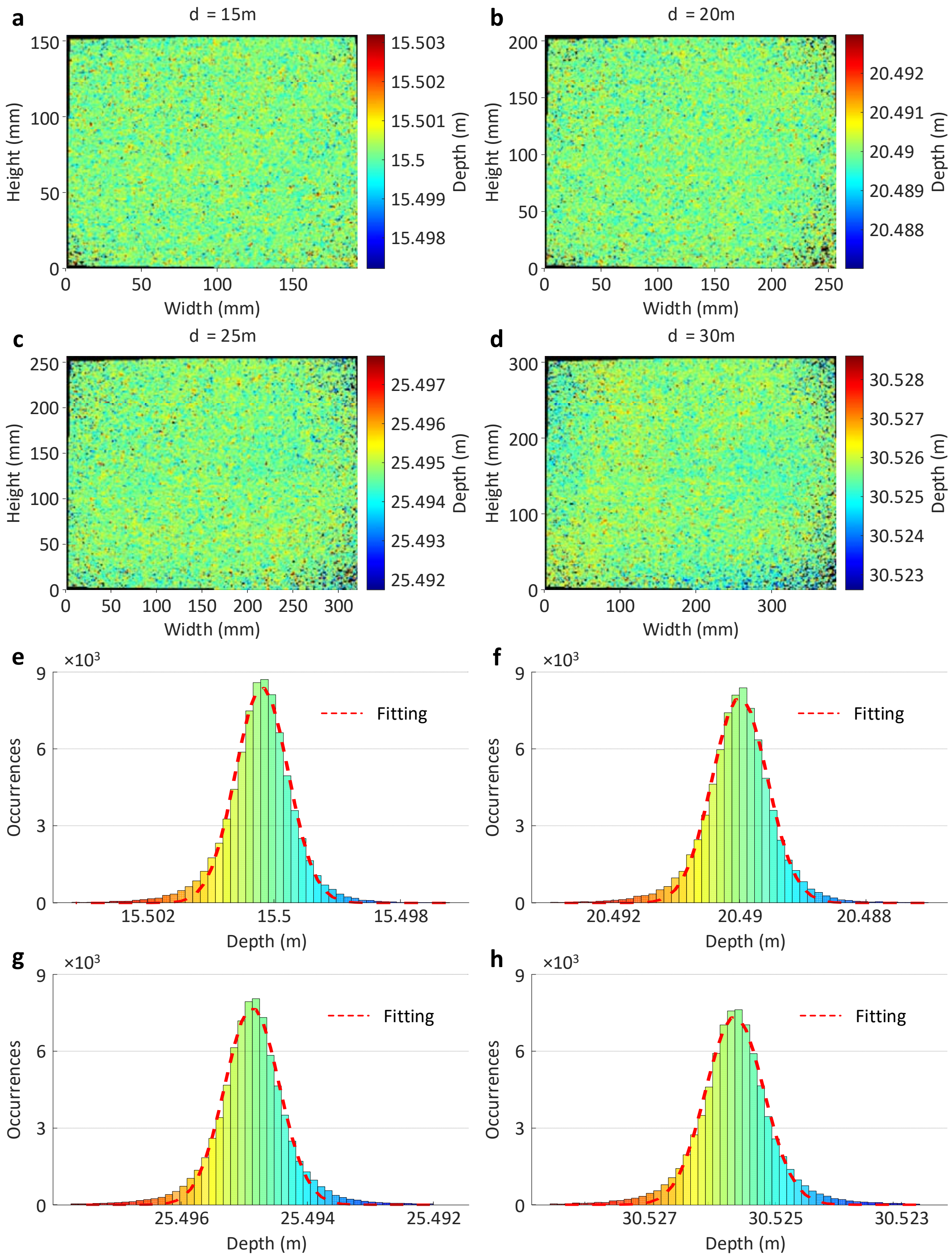

Fig. S4. Range precision characterization. a-d 3D images of the planar target at distances of 15 m, 20 m, 25 m, and 30 m. **e-h** Histograms corresponding to 3D images.

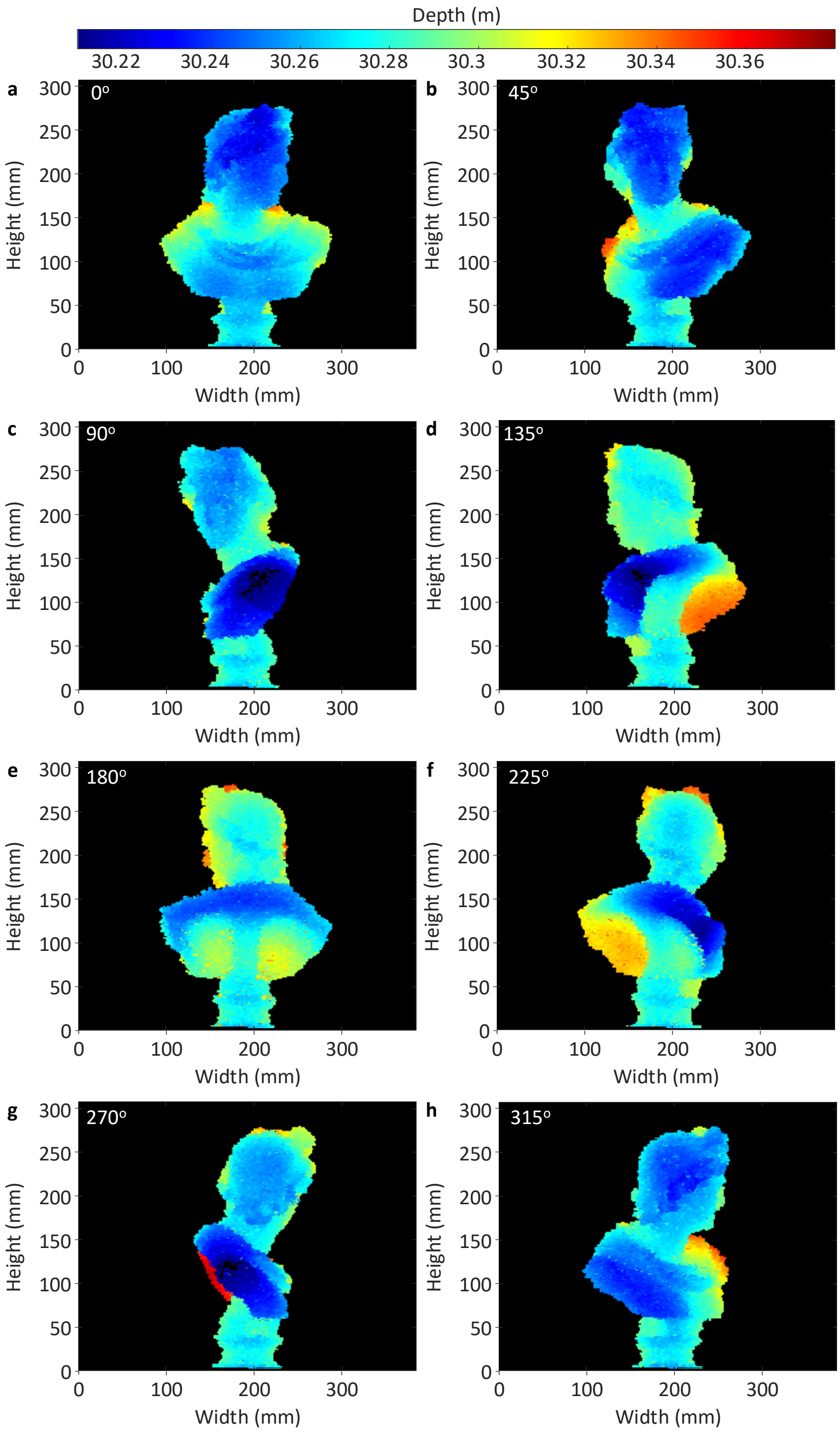

Fig. S5. 3D images of the bust sculpture at different orientations. **a**, 0°. **b**, 45°. **c**, 90°. **d**, 135°. **e**, 180°. **f**, 225°. **g**, 270°. **h**, 315°.

Parameters	Value
c	3×10^{8} m/s
Δf	1.97 MHz
N	3000
η	80%
h	6.63×10^{-34} J/s
T_e	4 ms
A_l	1.59 μV/m
A_e	31.94 nV/m
M	0.29
φ_{bias}	$\pi/2$
i_{dark}	11.86 fA
e	1.60×10^{-19} C
$Q_{n-readout}$	400 e^{-}

Table S1 Parameters of the range precision analysis.

Movie S1.

3D flower blooming. The color indicates the depth information.

Movie S2.

360-degree view of a bust sculpture. The color indicates the depth information at 0^{o} orientation.

Reference

S1. J. R. Janesick, T. Elliott, S. Collins, M. M. Blouke, J. Freeman, Scientific charge-coupled devices. *Opt. Eng.* **26**, 692–714 (1987).

S2. C. Nguyen, J. Park, Stepped-frequency radar sensors: theory, analysis and design (Springer International Publishing, 2016).

S3. S. M. Kay, Fundamentals of statistical signal processing: estimation theory (Prentice-Hall, 1993).

S4. D. Fink, Coherent detection signal-to-noise. *Appl. Opt.* **14**, 689–690 (1975).

S5. D. Dussault, P. Hoess, Noise performance comparison of ICCD with CCD and EMCCD cameras. *Infrared Systems and Photoelectronic Technology* (SPIE, 2004), **5563**, 195–204.