Coherence As Texture – Passive Textureless 3D Reconstruction by Self-interference

Supplementary Material

8. Relationship with OCT-based methods

Optical coherence tomography (OCT) [8] also extracts 3D information using interferometry. The main difference with our approach, however, is that OCT is an *active* approach.

OCT is a time-of-flight-based method, which measures the time delay caused by the path difference between the target and reference waves. Using reference waves means it often requires controlled light sources, often specialized lasers. One exception is the work of Kotwal et al. [10] who use OCT under sunlight. However, this isn't a fully passive approach because it captures the sunlight in a beam-splitter and redirects it into the target and a reference mirror. This direction of sunlight can disturb the subject. Also it tracks the position of the sun to direct it into the scene.

However, both OCT and our method are based on interferometry, and share a common constraint on *temporal coherence length*. Temporal coherence length is the maximal path length difference for a wave that can interfere with its delayed copy. For laser, it can be several meters long. But for a filtered white light source, the length is approximately $\lambda^2/FWHM$, where $FWHM$ means the bandwidth of the bandpass filter. In our case, the temporal coherence length is around $80\mu m$.

For OCT, temporal coherence length directly impacts the depth resolution. For our setup coherence length is less of an issue because the paths of our two interferometer arms have similar lengths. Still, when calibrating the self-interference setup, we need to make sure the length difference between two 4f systems is less than the temporal coherence length.

9. Range and resolution analysis

9.1. Simulation setting

To understand the resolution and range of our system in detail, we start with a numerical simulation.

We simulate a larger white plane placed at different depths from the setup, and fix the magnification ratio to $M = 1$. The reflected waves from the planes are composed of several coherent waves with wavelength $\lambda = 600nm$, each wave has a Rect support function with width Δc, and it has a uniform amplitude but random phases at resolution $0.25\mu m$.

We selected coherence and imaging parameters satisfying the contrast conditions of Eq. (13), $\Delta x = 0.75\mu m$, $\Delta\Phi = 2\mu m$ and $\Delta c = 16\mu m$. We set $\beta = 0.27$, which corresponds to a viewing angle of $15°$.To simulate an image of a plane in a certain depth, we iterate over all the coherent components of the wave and sum up their intensities. We use the Holotorch library in Python to simulate wave propagation.

9.2. Depth range

We start the derivation by considering the case of unit magnification $M = 1$ and adapt it to general magnification in Sec. 9.4.

As we discussed in Sec. 4.2, the range of depths we can measure is bounded because for far planes the interference pattern is too weak to be detected. Note that in the images I_t, we always measure the summation of the DC term and interference signal. Thus, while we can increase exposure or gain to amplify the interference amplitude, it will also magnify noise in the DC term. Therefore, we normalize the interference amplitude by the DC component of the observed images. That is, we define *contrast*, the strength of the measured signal, as:

$$C(x,y) \equiv \frac{2|\sum_t e^{j\phi_t} I_t(x,y)|}{\sum_t I_t(x,y)} = \frac{2|J(x,y)|}{K(x,y)} \tag{17}$$

where $K(x,y) = \sum_t I_t(x,y)$ is the DC component of the interferograms. For fully coherent waves u and v , the contrast term $C(x,y)$ equals $\frac{2|u||v|}{|u|^2+|v|^2}$. With partially incoherent waves we get weaker interference. Effectively, the contrast values are always between 0 and 1.

To understand what depth ranges we can cover we analyze the variance in the position of the peak x_p detected at each row (see Eq. (11)) and the contrast C_p at this pixel.

For each target depth we simulate 100 different planes at the same depth, and calculate the mean peak contrast $\bar{C}_p$, and plot it as a function of depth in Fig. 11(a). As expected, we can see that $\bar{C}_p$ decreases when the target plane is not focused.

We start with the relationship between depth and contrast, as summarized in the following claim.

Claim 3 *The mean peak contrast scales as* $\bar{C}_p = \frac{0.78}{\zeta}$, *where*

$$\zeta = \frac{\lambda z}{\Delta\Phi\Delta c} + 1 \tag{18}$$

is a "normalized depth".

We prove the result below, by combining a few supporting claims.

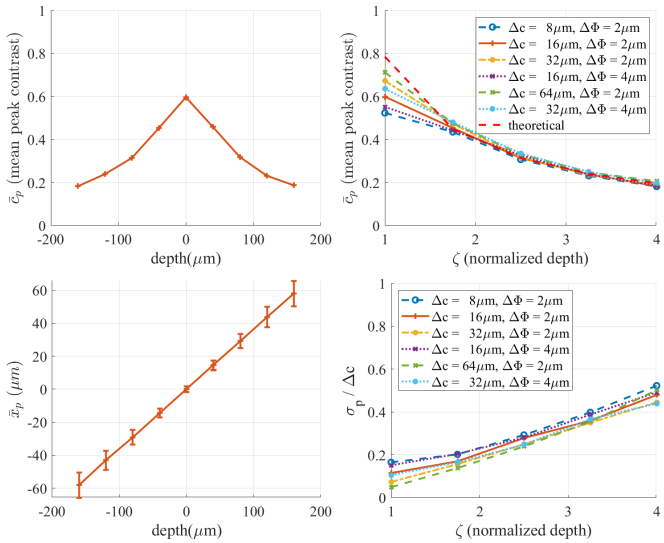


Figure 11. **Numerical analysis of range and resolution** (a) The contrast evaluated with the numerical simulation of Sec. 9.1, using planar targets at different depths. Contrast is highest when target is in focus (depth = 0), and decreases at larger distances. (b) Contrast for different Δc and $\Delta\Phi$ parameters, contrast is proportional to the normalized depth ζ. (c) The mean pixel position $\bar{x}_p$ is linear with depth. The standard deviation σ_p marked with bars around the curve, increases with defocused. (d) We evaluate σ_p for different experimental parameters and show that it scales with ζ.

The above claim not only shows the contrast is inversely proportional to the depth but also shows that the contrast is affected by Δc and $\Delta\Phi$. To verify the above claim, we repeat the same experiment using different values for Δc and $\Delta\Phi$. As shown in Fig. 11(b), all experiments demonstrate a consistent behavior agreeing with the theoretical value.

Empirically, for a reasonable detection, contrast should be larger than 0.2, which implies, following Claim 3, that we want $\zeta \leq 4$. A short calculation leads to the conclusion that we can measure objects inside the depth range:

$$|z| \leq \Omega_z, \quad \text{with} \quad \Omega_z = \frac{3\Delta\Phi\Delta c}{\lambda}, \tag{19}$$

which agrees with Eq. (14) of the main paper for the magnification $M = 1$.

To prove claim 3, we first note the reason for reduced contrast is the overlapping of multiple interference patterns. When deriving Eq. (10) from Eq. (9), we claim each wavefront $u^{n,m}$ has non-zero content only around $(n\Delta c, m\Delta c)$. However, when waves get defocused, their support spread and $u^{n,m}(x,y)$ can interfere with $u^{n,m}(-x,y)$ even when $n \neq 0$. When multiple interference terms overlap in the same sensor unit and each of them has a different phases, the overall contrast is reduced.

To caculate how the contrast is reduced, we start by discussing the support of the defocused wave as a function of aperture size.

Claim 4 *When $z \gg \Delta c$, the support of the defocused wavefront is $\frac{\lambda z}{\Delta\Phi}$.*

Proof: Consider one wavefront on the target $u^{n,m}(x,y)$ relayed to the input plane of the orthographic camera as in Fig. 4. It will first propagate distance z and then be constrained by an aperture of width D in the 4f system. Since only light rays whose angle is within the range $\frac{D}{f}$ pass, by simple ray optics considerations, the support of the "defocus blur", namely the sensor area at which rays passing through the aperture hit the sensor, is $\frac{zD}{f}$. Note that the aperture shift b will shift the defocus blur position on the sensor but will not change its width. The diffraction blur kernel equals $\Delta\Phi = \frac{\lambda f}{D}$. Therefore, we can rewrite the support as $\frac{\lambda z}{\Delta\Phi}$.$\square$

Claim 5 *The number of incoherent components interfering in each sensor point is ζ^2, with*

$$\zeta = \frac{zD}{f\Delta c} + 1. \tag{20}$$

Note that the term ζ is a linear function of the depth z, hence we refer to it as the "normalized depth". *Proof:* Since the center of the waves $u^{n,m}(x,y)$, $u^{n,m}(-x,y)$ are separated horizontally by $2|n|\Delta c$, they will interfere if $2|n|\Delta c < \frac{\lambda z}{\Delta\Phi}$. For example, if $\frac{\lambda z}{\Delta\Phi\Delta c} = 2$, then interferences occur for n values $n = -1, 0, 1$. Thus the number of incoherent components interfering is ζ^2, with $\zeta = \frac{zD}{f\Delta c} + 1$. $\square$

Next we derive how the number of interfered components effects the contrast we can measure. For this, we review a standard result in statistics, showing that with N independent coherent components the contrast scales as $\frac{1}{\sqrt{N}}$.

Claim 6 *Consider N pairs of independent random variables $U_1, \ldots U_N$, $V_1, \ldots V_N$, then*

$$\frac{E\left[\sum_n U_n V_n^*\right]}{E\left[\sum_n |U_n|^2 + |V_n|^2\right]} = \frac{1}{\sqrt{N}} \frac{E\left[U_1 V_1^*\right]}{E\left[|U_1|^2 + |V_1|^2\right]} \tag{21}$$

The intuition behind this result is that the numerator of the contrast is the summation of N independent complex values, while the denominator sums N positive values. When complex values are summed, terms can cancel each other and reduce contrast.

By combining claims 5 and 6 we see that since we average ζ^2 number of waves, the contrast is inverse proportional to $\frac{1}{\zeta}$. Finally, if both real and imaginary parts of U_n and V_n are Gaussian variables with the same variance, the expected contrast is around 0.78 when $N = 1$. By combining these arguments we arrive at claim 3.

9.3. Depth resolution

As derived in Eq. (12), the estimated depth is $\hat{z} = \frac{x_p}{\beta}$. Therefore, the resolution at which we can detect depth depends on the accuracy at which we can detect x_p. As illustrated in Fig. 7, the interference pattern we image is a speckle pattern whose width is a few pixels, and the detected maximal x_p can somewhat vary inside the speckle pattern. We define σ_p to be the standard deviation of the x_p position. We numerically compute this variance using the numerical simulation described in Sec. 9.1, by sampling multiple random realizations for each depth plane. σ_p are plotted in Fig. 11(c), demonstrating that the standard deviation of the detected depth increases when we are further from the focal plane and wider defocus blur is present. In Fig. 11(d), we repeat the simulation for a few other imaging configurations and observe that the standard deviation is proportional to the normalized depth ζ of Eq. (18). As mentioned above, in practice we can detect depth in the range $\zeta \leq 4$. Within that range we empirically observe that the average σ_p value, is around around $0.3\Delta c$. Since the depth is $\beta^{-1}x_p$, this leads to the conclusion that the depth resolution is

$$\Delta z \approx \frac{0.3\Delta c}{\beta}. \tag{22}$$

While increasing the tilt angle β improves depth resolution, in practice wide angles are more susceptible to optical aberrations.

9.4. Magnification

As stated in Sec. 3.3, before the self-interference part, we can add an additional lens to scan and scale the scene. Below we derive how such lens magnification changes the depth range and resolution that we can recover. We show that the range and resolution are scaled linearly with the magnification, but the number of distinguishable depth planes does not change.

To see this consider Fig. 12. A lens magnifying the target by a factor M will have two effects. First the spatial size of features is scaled by M, and in particular, if the coherence length of the illumination hitting the actual target is Δc, the coherence length of the scene imaged by this lens is $M\Delta c$. On the other hand, the depth planes are scaled by M^2.

The fact the depth is scaled by M^2 means that if without the magnifying lens the our system could cover depth range Ω_z, the depth ranged mapped into this range by the magnifying lens is Ω_z/M^2.

One the other hand, the depth range and resolution derived in Eqs. (14) and (22) depend on Δc, and Δc is scaled by M. As a result, the depth range we can cover is only Ω_z/M.

A similar argument shows that the depth resolution Δz is scaled to $\Delta z/M$.

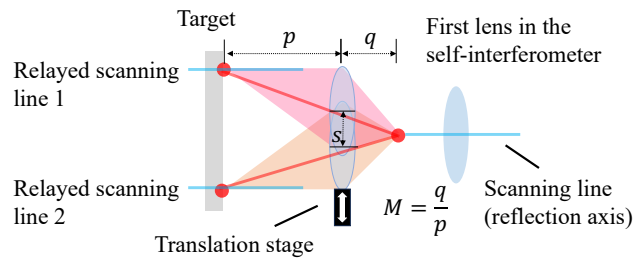


Figure 12. **Translating an additional lens to scan the scene.** Since our main setup can only scan a vertical slice of the scene, we add a relay lens in front of the main setup. By translating this lens we can scan different lines of the target, since a different strip of the scene is mapped to the reflection axis of the main setup. The relay lens can also magnify the scene.

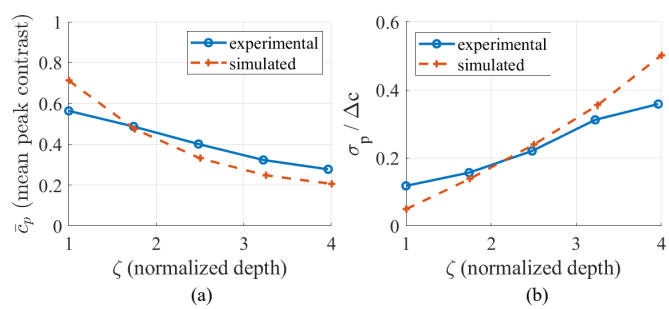
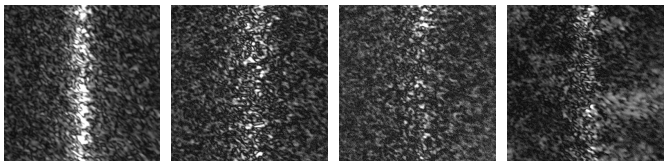


Figure 13. **Comparison between real and simulated results.** (a) Comparing contrast. (b) Comparing the standard deviation of peak position σ_p. While we do not know precisely all the parameters of the real system, the measured and simulated curves follow a similar behavior.

Since both Ω_z and Δz are scaled by the same factor, the number of depth planes we can distinguish does not change with magnification. However, note that $M\Delta c$ needs to remain larger than $\Delta\Phi$, so we cannot scale the scene arbitrarily small.

9.5. Comparing numerical simulation to experimental measurements

In Fig. 13 we compare the contrast and the variance of the mean peak position between numerical simulations. For that we use the experiment described in Sec. 5.2 and Fig. 8 of the main paper, were we vary the position of a planar target on a motorized stage and attempt to estimate the depth of these images. We compare the variance and contrast of the real system to the ones predicted by our analysis and numerical simulation above. While some differences exist, both real measurements and numerical predictions follow a similar behavior.



(a) Aluminum (b) Paper (c) Resin (d) Foam

Figure 14. **Interference images with different materials.** While a metallic target results in a strong, narrow interference image, materials with more subsurface scattering result in interference images of wider support, hence depth estimation is noisier.

10. Results with different materials

10.1. Materials with subsurface scattering

As mentioned in the main paper, subsurface scattering blurs the wavefront and hence reduce the interference contrast. To minimize this problem, most of the results in the main paper use metallic targets.

Here we test the effect of subsurface scattering. In Fig. 14 we test the coherence of a few material. We capture plane targets under a swept-angle illumination with $\Delta c = 54\mu m$ and $\beta \approx 0.05$. For the aluminum target in Fig. 14(a), the interference image is a clear vertical line. However, the line has a wider spread for the paper and resin targets. For the foam target, interference noise can appear other the entire frame. Below we show that the wider support reduces depth resolution.

10.2. Results with swept-angle illumination

In Fig. 15 we present additional depth acquisition results with a few other targets, and with different materials. In this figure we use the monochromatic swept-angle illumination. The first row is a metal statue of an old man. As the target is metallic, we can reconstruct details such as the height difference between nose and beard. The second row is a paper plane target. While the paper has subsurface scattering, we can still reconstruct depth planes since the target is rather simple. The third row is a resin statue of "the thinker". The reconstruction is recognizable, though some subsurface scattering reduces the reconstruction accuracy. The last row is a resin cat with painted eyes and mouths. The specularity in the eye and the texture change near the mouth, leading to reconstruction artifacts.

For non-metallic targets under sunlight, the contrast reduction now combines two factors, subsurface scattering and the non-monochromatic illumination. Overall the interference contrast is too weak and we did not manage to reconstruct such targets.

11. Reconstruction algorithm details

The naive depth extraction equation described in Sec. 3.2, is to find the peak of the interference amplitude in each row and calculate the target depth from the shift.

In practice, processing each row independently is very noisy, because the interference signal contains speckles. To improve robustness we use the following filtering stages, illustrated in Fig. 16.

First, since image intensities may not be uniform, we normalize the interference amplitude by the DC component before extracting its peaks.

Second, we assume the target is smooth and blur the interference signal with a 2D Gaussian filter before extracting its peak. We blur the vertical axis with a Gaussian of s.t.d. 50 pixels, resulting in a similar depth estimate in nearby rows. We blur horizontally with a smaller s.t.d of 15 pixels to eliminate some of the speckles. The extracted peaks after blurring are visualized in Fig. 16(d).

In the third stage, we further eliminate noise by using the Viterbi algorithm to select the peak of each row while forcing nearby rows to have similar values. The result of this stage is visualized in Fig. 16(e).

Finally, after we combine multiple vertical scans, we further smooth the depth map by applying a small horizontal Gaussian filter of s.t.d 2 scans (corresponds to $\approx 200\mu m$).

12. Prototype detail

12.1. Michelson interferometer with LC cell

As mentioned in Sec. 5.1, for better stability we use an LC cell to delay one arm instead of translating a mirror, (translating a mirror with sub-wavelength accuracy requires very precises stages).

Since an LC cell is a birefringent component that can delay linearly polarized light aligned with its fast axis, if the two paths of the interferometer have orthogonal polarizations, only one arm is delayed. To do this, as shown in Fig. 17, we first linearly polarized the light at 45 degrees to ensure horizontal and vertical polarized light are coherent. Thereafter, we put two linear polarizers in the two paths, one with oriented vertically and the other is oriented horizontally. We align the fast axis of the LC cell with one path and thus only delay that path. Finally, we need another linear polarizer rotated by 45 degrees to combine the two orthogonal paths and interfere them.

12.2. Shifting lens for scanning

In Sec. 3.3, we mentioned that we scan the scene by shifting a lens in front of the setup. We further explain this in Fig. 12. This shift makes different vertical lines from the scene mapped to the flipping axis of the main system; hence, effectively, we flip along different lines in the scene.

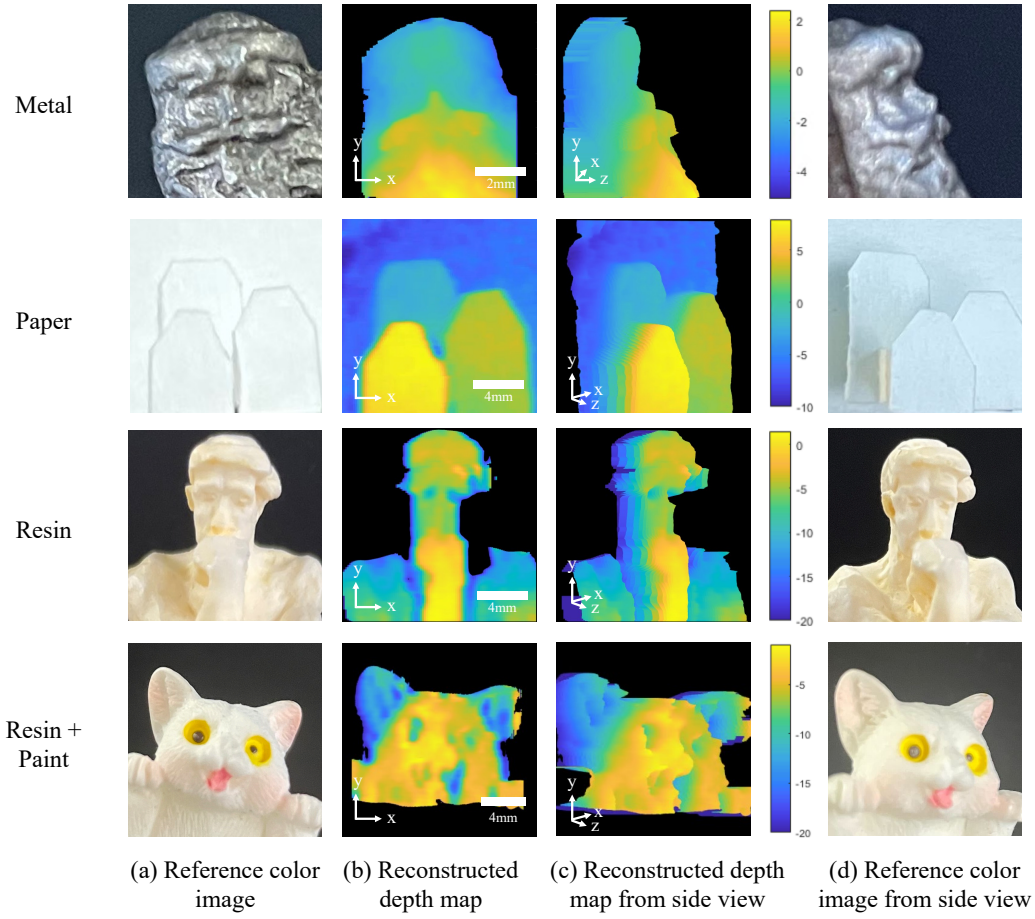
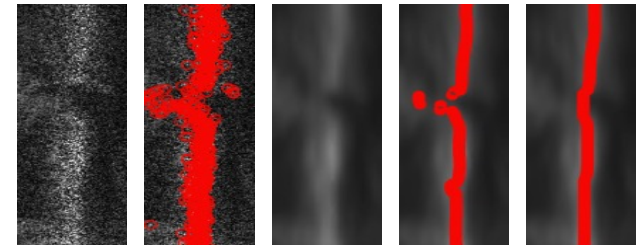


Figure 15. **Reconstruction results with swept-angle light source.** Targets in different rows are made of different materials. Targets with stronger subsurface scattering (paper, resin) have reduced depth resolution and can have artifacts when texture or material changes.



(a) Contrast (normalized interference image) (b) Extracted peak (c) Filtered contrast image (d) Extracted peak from filtered image (e) Extracted peak with Viterbi

Figure 16. **Steps of the peak detection algorithm.** (a) The captured interference image, dominated by speckle noise. (b) Detecting the peak position in every row independently leads to noisy results. (c) Filtering the interference image (d) Peaks detected from the filtered images are smoother. (e) We further improve the depth extraction using the Viterbi algorithm.

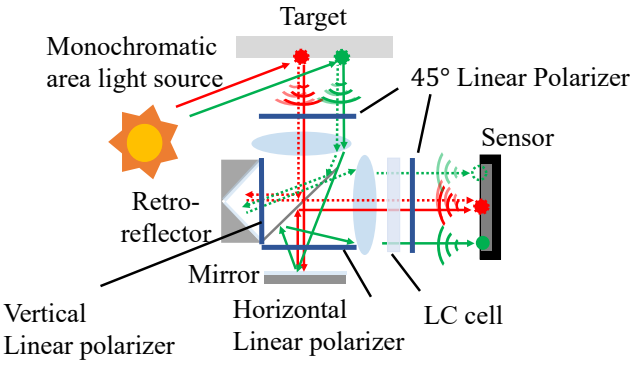
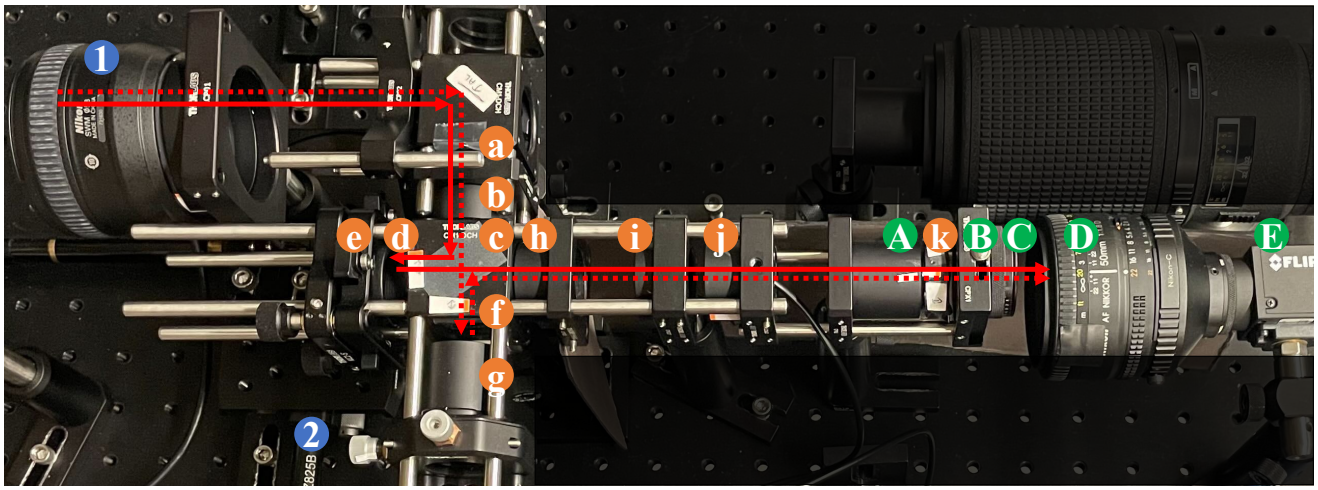


Figure 17. **Polarized interferometer.** Our interferometer is implemented using polarization rather than a translating mirror. The two arms are designed to have orthogonal polarization and used an LC cell to delay only the horizontally polarized waves.



1 Lens (f = 50mm)
2 Translation stage
a 45° linear polarizer
b Lens (f = 75mm)
c Beam splitter
d Horizontal linear polarizer
e Mirror
f Vertical linear polarizer
g Hollow roof retroreflector
h Lens (f = 75mm)
i Bandpass filter (Thorlab FLH633-5)
j LC cell (Thorlab LCC1115-B)
k 45° linear polarizer
A Lens (f = 50mm)
B Translation cage plate (Thorlab CPX1)
C Aperture
D Lens (f = 50mm)
E Camera (FLIR BFS-U3-120S4M-CS)

Figure 18. **Prototype.** List of components used in our setup.

Also, the amount of shifting lens depends on the range of the target to scan as well as the magnification rate M. For a target width w, the lens needs to be shifted by $w/(1/M + 1)$, and we mount the lens on a motorized translation stage to shift the lens in around $0.1mm$ resolutions.

12.3. Components list

Following the schematic of Fig. 5, we implement a hardware prototype as in Fig. 18. For the self-interferometer marked in orange, we use two lenses with $f = 75mm$. Both the mirror and the hollow roof retroreflector are placed at the Fourier plane. As discussed in Sec. 12.1, we use an LC cell with several polarizers to replace the translation stage.

For the tilted orthographic camera system marked in green, we use two lenses with $f = 50mm$. The aperture in the Fourier has an adjustable size D and a controlled horizontal displacement b. The camera sensor has pixel pitch $\Delta x = 1.85\mu m$. For the scanning lens marked in blue, we use a camera lens with $f = 50mm$ mounted on a motorized translation stage. It can imaged a target at $2f = 100mm$ away with magnification ratio $M = 1$ and $3f = 150mm$ away with $M = 0.5$.

12.4. Calibration detail

In this section, we describe the steps we use to build and calibrate our setup in detail. We suggest the reader first prepare a swept-angle light source [9], which has an adjustable coherence length. The rest of the steps are as follows:

1. Mount the beamsplitter (c). The whole set-up will be built around the beamsplitter.
2. Mount the mirror (e) on a kinetic mount that can adjust the tilt angle, and attach it to the beamsplitter (c).
3. Attach one lens (b) to the beamsplitter. Calibrate its axial position so the mirror is f-away(75mm) from the lens.
4. Mount the hollow roof retro-reflector (g) on a translation stage that can adjust lateral positions and attach to the beamsplitter. Calibrate the axial position of (g), so it is also f-away(75mm) from the lens (b).
5. Temporarily put a target f-away (75 mm) behind the lens (b). Observe it from the camera focused at infinity (D,E). We should see that the target and its flipped version are both in focus.
6. Adjust the angle of the mirror (e), so the center of the target is aligned with its flipped version.
7. Attach the second lens (h) to the beamsplitter and calibrate it to be f-away(75mm) from both the mirror (e) and the retro-reflector (g).
8. Attach another lens (A) behind the lens (h). The distance between them is a summation of their focal lengths

(75mm+50mm), so the camera focused at infinity (D,E) can again see the target in focus.

9. Mount the aperture (C) on the translation cage plate (B) and calibrate it to be f-away(50mm) from the lens (A). We first keep (B) in the center position.
10. Attach cross polarizers (d,f) onto the beam splitter. We suggest slightly slanting the polarizer in vertical directions to avoid ghosting.
11. Mount 45-degree linear polarizers (a,k) as well as LC cell (j).
12. Illuminate the target with the swept-angle light source. When performing phase-shifting interferometry with LC cell (j), we should be able to observe a high contrast in the reflection axis as in Fig. 8.
13. Calibrate the lateral position of the retro-reflector to maximize the contrast.
14. Shift the aperture on the translation cage plate (B,C); now, the targets at different distances will result in different contrast peak positions.
15. Add the bandpass filter (i) to enable using a white light source. We also need to use a motorized translation stage to finetune the axial position of the retro-reflector (g), so the difference between the optical length of two arms is less than $80\mu m$.
16. Mount an additional lens (1) on a translation stage (2) to enable scanning the full scene.