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

Wei-Yu Chen[1]*, Aswin C. Sankaranarayanan[1], Anat Levin[2], Matthew O'Toole[1]
[1]Carnegie Mellon University, [2]Technion – Israel Institute of Technology

Abstract

Passive depth estimation based on stereo or defocus relies on the presence of the texture on an object to resolve its depth. Hence, recovering the depth of a textureless object—for example, a large white wall—is not just hard but perhaps even impossible. Or is it? We show that spatial coherence, a property of natural light sources, can be used to resolve the depth of a scene point even when it is textureless. Our approach relies on the idea that natural light scattered off a scene point is locally coherent with itself, while incoherent with the light scattered from other surface points; we use this insight to design an optical setup that uses self-interference as a texture feature for estimating depth. Our lab prototype is capable of resolving depths of textureless objects in sunlight as well as indoor lights.

1. Introduction

Shape recovery is a problem of fundamental importance in computer vision, where the goal is to recover a 3D description of a scene from one or more images. In passive settings, shape information can be recovered from disparity [13], shading [23, 24], focus [7], defocus [16], motion [15], or even polarization [19, 22]. All of these approaches, however, rely on making certain assumptions about the scene, limiting the generality of each approach.

Consider, for example, the scenario shown in Figure 1(b), where a camera measures a textureless surface lit by an unknown white light source. Upon first glance, the 3D reconstruction problem appears to be underconstrained. It has been widely accepted that it is impossible to passively reconstruct the depth of such a textureless plane through stereo imaging or depth from (de)focus; irrespective of the viewing angle and focus settings of the camera, the captured images will always be uniform in brightness, providing no visual information that can be used for 3D reconstruction.

So when is it possible to see the shape of a textureless object, under unknown illumination? The classic aperture problem [14] tells us that correspondences cannot be found for textureless objects and hence, disparity across viewpoints is unobservable. Shape from shading [23, 24] and shape from polarization [19, 22] permit the recovery of the normal of a textureless plane, but not its depth. Sundaram et al. [20] conclude that the depth of a textureless plane can be reconstructed, but only if the plane is heavily tilted with respect to the camera viewpoint. To overcome the lack of texture, numerous works apply active imaging approaches such as structured light [2] or time-of-flight cameras [5, 8]. Kotwal et al. [10] have demonstrated that OCT based depth acquisition can be implemented under sunlight without coherent laser illumination, yet their approach is not fully passive because it uses optics to control and direct the sunlight illumination hitting the scene. The main problem of active

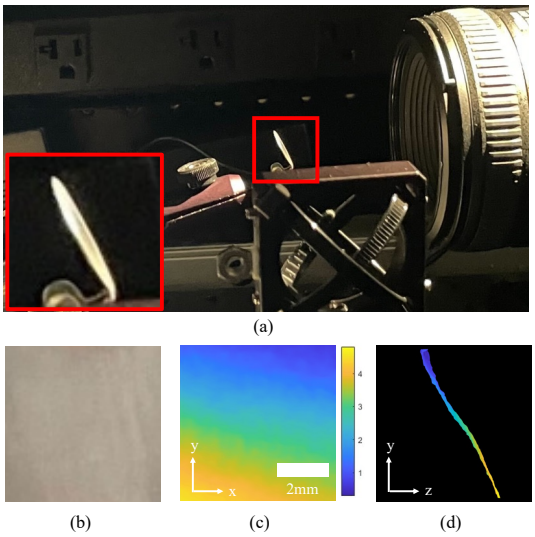


Figure 1. **Passive 3D reconstruction of a textureless plane under white, incoherent illumination.** (a) A photo of the capture setup. A tilted plane target (enlarged in the inset) lies in front of the scanning lens of our setup. (b) Front view of the target. Within the sensor area, it is a uniform, textureless plane target. (c) Depth estimation with our approach, resolving the desired tilt. Note that we can measure the explicit depth, which can not be recovered by other passive methods that only measure normals. (d) A side view of the reconstructed target reveals it is a tilted plane.

*Corresponding author: wyharveychen@gmail.com

approaches is that projecting illumination may not be applicable under strong ambient light conditions such as sunlight. They are also inapplicable in many live imaging scenarios where projected illumination can disturb the subject.

In this work, we demonstrate for the first time a passive reconstruction of the depth of textureless subjects, by using the *local coherence properties of the light as a form of texture*. The key idea is to use coherence to solve a correspondence problem, even on a uniform target. We then measure the disparity shift of such points and convert it to depth. To this end, we build an interferometric setup that interferes the light emerging from a scene with a mirror flipped version. Under natural illumination which is spatially incoherent, only points on a central column of the image will interfere with the flipped copy, since their location remains the same after the flip. That is, every point other than the centered one is masked out. Since only one scene point in a row is visible, we can measure its disparity shift and convert this shift to depth.

In the following sections we start by reviewing the principle of self-interferometry under natural illumination, explain our optical setup which measures interference with a flipped copy and our resulting depth extraction pipeline. We analyze the range and resolution supported by our device. Finally we demonstrate a working prototype and its resulting 3D capture. We also release our simulation code and prototype details on the project website.[1]

Contributions. This work proposes to passively reconstruct shape without relying on texture or shading cues, offering the following contributions:

- *Coherence As Texture.* With a self-interference version of a Michelson interferometer, we show that coherence can be used to estimate depth of a scene point illuminated with natural light.
- *Depth range and resolution analysis.* We carefully analyze the range and resolution supported by our system, and validate the theoretical derivation using numerical simulations.
- *Lab prototype.* We build a lab prototype that can 3D reconstruct textureless targets and demonstrate our setup can passively work under incoherent natural illumination.

Limitations. The success of using coherence as a texture critically depends on the coherence length of the illuminating light source. While coherence length of sunlight, for example, is ample for robust 3D shape recovery, many extended light sources have coherence lengths that are too small to lead to successful depth scans. We analyze this in detail. Our method is also susceptible to sub-surface scattering as it tends to increase incoherence. Our system requires long acquisitions times due to the use of a narrowband spectral filter that blocks most of the incident light as well as the need to scan the scene one 'column' at a time.

[1] https://imaging.cs.cmu.edu/coherenceastexture

2. Background

2.1. Review of interferometry

To understand interferometry, we start by considering a classical Michelson interferometer setup as shown in Fig. 2(a). A coherent light is emitted from a laser source and is split into two copies via a beamsplitter. One copy illuminates a target scene of interest and the other reflects from a mirror. The two returning wavefronts recombine at the beamsplitter and are measured by the camera. Denoting the waves returning from the two arms by $u(x,y), v(x,y)$ respectively, and assuming the two waves are fully coherent with each other, we can express the measured camera intensity as

$$\begin{aligned} I(x,y) &= |u(x,y)+v(x,y)|^2 \\ &= |u(x,y)|^2+|v(x,y)|^2+2Re(u(x,y)v(x,y)^*). \end{aligned} \tag{1}$$

This measured intensity is the sum of the individual intensities of each of the waves $|u(x,y)|^2+|v(x,y)|^2$, plus the real part of an *interference* signal

$$J(x,y) \equiv u(x,y)v(x,y)^*. \tag{2}$$

We can extract the interference signal using phase-shifting interferometry (PSI) [6]. For that, we add a phase shift to one of the wavefronts, e.g., by slightly translating the mirror to modify the optical path length of v, so that we measure:

$$I_t(x,y) = |u(x,y)+e^{j\phi_t}v(x,y)|^2, \tag{3}$$

using $T \geq 3$ equally spaced phase delays $\phi_t = 2\pi(t-1)/T$ we can extract

$$J(x,y) = \frac{1}{T}\sum_t e^{j\phi_t} I_t(x,y) = u(x,y)v(x,y)^*. \tag{4}$$

Note that we can determine whether u and v are really coherent from the power of the interference term $|J|$. For fully coherent waves, the power of the interference term can be as high as $|J| = |u|\,|v|$. On the other hand, if u and v are fully incoherent wavefronts, e.g. they are generated by two independent laser sources as in Fig. 2(b), no interference will be measured and $|J| = 0$.

2.2. Self-interference under natural illumination

As part of this work, we measure interference under natural illumination. Unlike the fully coherent lasers analyzed above, natural illumination is spatially incoherent. Below we model spatial incoherence and explain what interference

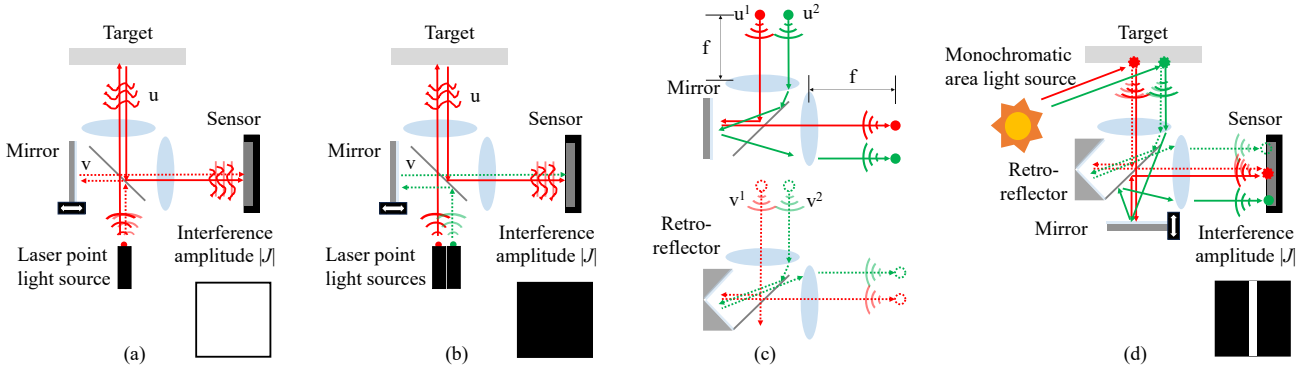


Figure 2. **From Michelson interferometer to our incoherent texture isolation.** (a) A classical Michelson interferometer measuring the interference signal between two waves. (b) Two incoherent waves, e.g., waves generated by two different sources, lead to zero interference signal. (c) A 4f system imaging a target scene with a mirror in the Fourier plane (upper part) and with a retro-reflector in the Fourier plane (lower part). The retro reflector results in a flipped image. (d) Combining the two parts of (c) as two arms of an interferometer, we self interfere the image with a flipped copy. Therefore, we get non-zero interference only for on axis points, located on the central column of the image. This coherence is generating "texture" even on a uniform target.

can be measured. Throughout this paper, we place a narrow-band filter in front of the camera so that the measured light is relatively monochromatic, even if it is spatially incoherent.

We express the wavefront illuminating the scene as a sum of mutually incoherent wavefronts

$$u(x,y) = \sum_{n,m} u^{n,m}(x,y), \tag{5}$$

and each wavefront has non zero power only at a local window around the point

$$(x^n, y^m) = (n\Delta c, m\Delta c), \tag{6}$$

where Δc denotes the coherence length of the illumination. We assume that the other copy of the interferometer is passing through some optical modulation g (to be defined below), resulting in a wavefront $v(x,y) = \sum_{n,m} v^{n,m}(x,y)$ with $v^{n,m} = g(u^{n,m})$. While $u^{n,m}$ are incoherent with each other, there is still coherence between each $u^{n,m}$ to each $v^{n,m}$, thus a Michelson interferometer will measure

$$I_t(x,y) = \sum_{n,m} |u^{n,m}(x,y) + e^{j\phi_t} v^{n,m}(x,y)|^2, \tag{7}$$

and with PSI we can isolate

$$J(x,y) = \sum_{n,m} u^{n,m}(x,y) v^{n,m}(x,y)^*. \tag{8}$$

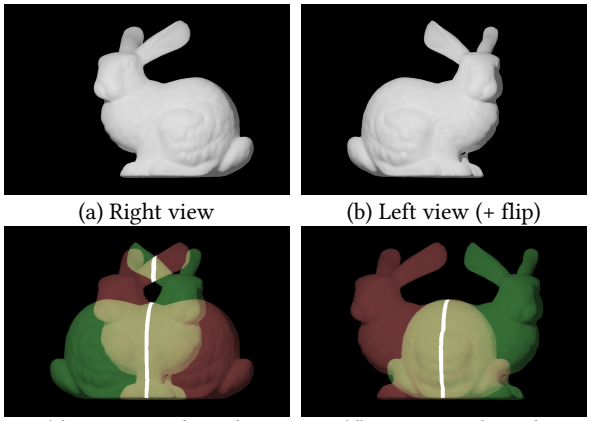
This approach is known as self-interference [18]. Previous approaches such as FINCH [1, 17] and COACH [21] use self-interference to capture a convolution of the scene with a complex point-spread-function (PSF), which is then used to extract 3D information. However, these approaches are not useful in textureless scenes since a uniform scene convolved with any PSF is still uniform.

3. Coherence As Texture

One of the more established approaches for depth estimation is stereo vision, where a target scene is imaged from two distinct viewpoints. This is typically done by capturing two images, and computing correspondences between the two images. Through a triangulation procedure, one can then use stereo correspondences to recover a depth map of the scene. However, traditional stereo matching largely relies on the existence of texture in the scene.

In this work, we propose an optical solution to the correspondence problem, one that relies on the coherence properties of the scene itself. Under natural illumination, our key observation is that the light emanating from a point on an object can interfere with itself. In contrast, merging the light emanating from two different points does not produce an interference signal. The central idea behind this work is to use this as an indicator function for computing correspondence. In effect, we use coherence as a texture.

Our approach starts with forming a pair of images captured from different perspectives (Figures 3(a-b)), with one image flipped along the horizontal direction. Provided sufficient textures, these images could be used directly in a conventional stereo procedure to recover depth, after unflipping the one image. Instead, our proposed approach is to optically interfere these two images (Figures 3(c)). Interference only occurs when a scene point maps light to the same camera pixel, which happens when this point lies on the reflection axis. The position on the camera sensor is related to disparity and can be used to infer the geometry of a cross-section of the target. To scan the entire object, one can laterally shift these two images in opposite directions, namely flip along a different vertical line, effectively performing a push broom scan of the scene (Figures 3(d)).



(a) Right view (b) Left view (+ flip)

(c) Superposed result (d) Superposed result (with shift)

Figure 3. **High-level overview of the proposed approach.** (a) Right image from stereo pair. (b) Left image from stereo pair, with an additional horizontal flip. (c) When superposing both images (shown in red and green), a subset of camera pixels observes light from the same 3D point (shown in white). We identify these pixels through optical interferometry, and use the pixel coordinate to infer disparity and scene depth. (d) The shape of the entire image is captured by optically shifting the two images and repeating the interfereometric process.

The remainder of this section presents a theory of depth estimation using coherence as a proxy for texture, and provides a design of an optical system to measure depth of textureless scenes.

3.1. Isolating coherent components

Even when imaging a uniform intensity scene, under natural illumination the scene can be expressed as a sum of local mutually incoherent wavefronts as described in Sec. 2.2. Our goal below is to use self interference to separate these different wavefronts and use them to obtain some localized texture, which is in turn used to extract depth information.

Our interferometric setup consists of two arms, with a mirror in one, and a retroreflector in the other. Figure 2(c) visualizes these two arms separately. The first arm, consisting of the mirror, simply images the scene through two lens that are positioned to form a 4f system between the target and the sensor. The mirror is placed at the Fourier plane of this system. As a result, the wavefront of the target $u(x, y)$ is formed on the sensor by this arm. In the second arm, we use a hollow-roof retro-reflector in the Fourier plane of the 4f system, which flips a wavefront around the vertical axis, thereby generating a flipped wavefront $v(x, y) = u(-x, y)$ at the sensor. Figure 2(d) shows the individual arms in Fig. 2(c) combined to form an interferometer. This setup was previously used by [9] to visualize the coherence property of a light source; here, we use it to isolate coherent components of the wavefront.

Following Eq. (8), the Michelson interferometer in Fig. 2(d) will measure the interference between the local incoherent components of u and v:

$$\begin{aligned} J(x,y) &= \sum_{n,m} u^{n,m}(x,y)v^{n,m}(x,y)^* \\ &= \sum_{n,m} u^{n,m}(x,y)u^{n,m}(-x,y)^*. \end{aligned} \tag{9}$$

Since each wavefront $u^{n,m}$ has non zero content only around $(n\Delta c, m\Delta c)$ (see Eq. (6)) all off-axis wavefronts are canceled from the above summation and we are left only with

$$J(x,y) = \sum_{m} u^{0,m}(x,y)u^{0,m}(-x,y)^*. \tag{10}$$

In other words, since the individual wavefronts are only coherent with themselves, if we self-interfere the waves with flipped copies, we mask out every wave which does not lie on the $x = 0$ axis. To see this, consider Fig. 2(d). Two rays emerging from the central point marked in red, are mapped by the two arms of the interferometer to a single point on the sensor, hence they can interfere. The two rays emitted by the off-axis green point are mapped by the two arms to two different sensor points, since the flipped copy is in a different position. These two far copies do not interfere.

We note that along each row of the interference image $J(x, y)$, there is a single non zero wavefront. Once this wavefront is isolated we can measure its disparity shift, as described below, resulting in a depth estimate.

3.2. Depth from disparity shift

With the interferometer described above only the central point of each row is measured. The next part of our system is designed to measure *disparity shift*, namely, the projection of each point shifts along the corresponding row and the displacement is proportional to its depth.

To this end we want an orthographic projection of the line pattern, at a tilted angle. We can build such an orthographic camera by putting a small aperture in the Fourier plane of a 4f system. To control the tilting angle we shift the aperture as in Fig. 4, so only rays in a tilted angle can be imaged. To be precise, when the aperture is shifted b away from the center, the orthographic camera only accepts light rays tilted at angle $\tan^{-1}\beta$ where $\beta = b/f$. Thus, when the target point is at distance z from the plane of focus, its projection will be shifted by βz.

Estimating depth with this measurement is straightforward. For each row we find the peak position

$$x_p(y) = \text{argmax}_x(|J(x,y)|), \tag{11}$$

and then predict

$$\hat{z}(y) = \frac{x_p(y)}{\beta}. \tag{12}$$

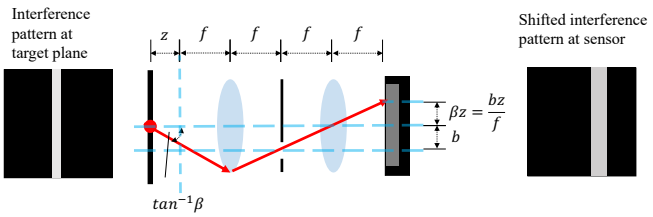


Figure 4. **Disparity shift with a tilted orthographic camera.** We build a tilted orthographic camera by placing a shifted aperture at the Fourier plane of a 4f system. The tilted orthographic camera only accepts rays at a tilted angle, and the resulting image shifts horizontally. This shift is a linear function of depth.

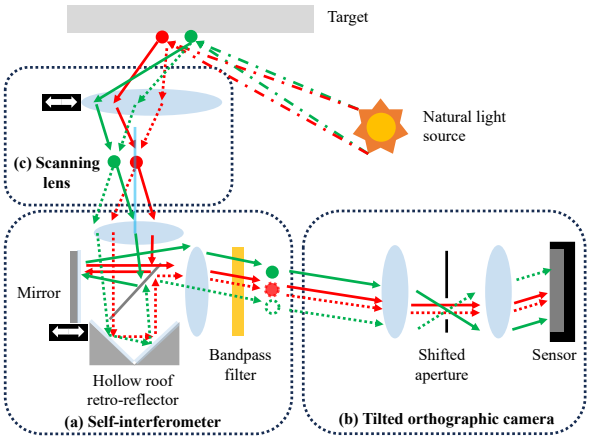


Figure 5. **Full setup,** combining the ideas of Fig. 2 (d), and Fig. 4.

3.3. Scanning and scaling the target

So far we have seen how to reconstruct the depth of one vertical line on the scene. To measure a wide target, we need to translate the setup horizontally to scan the entire width of the scene. To simplify scanning, we introduce a scanning/imaging lens that brings the scene onto focus in front of the 4f relay, as shown in part (c) of Fig. 5. Instead of translating the entire setup, we chose to shift just this lens; as we show in the supplemental material, this is sufficient to scan the width of the entire target scene. The same lens also allows us to magnify the scene.

3.4. Full 3D reconstruction

We combine the components of the setup described in the previous subsections into a full 3D acquisition system illustrated in Fig. 5. More details on the setup are provided in supplement.

In Fig. 6 we illustrate our full reconstruction pipeline. The target has a stair structure, each stair has a different depth, but almost uniform brightness. Despite the fact that the images I_t appear uniform, by capturing 4 different I_t images corresponding to different shifts of the mirror in the Fourier plane of the interferometer, we can extract the interference component J in Fig. 6(d). As predicted above, for each row of the image J there is only one region with strong interference. The shift of the stripe corresponds to depth. We extract the peaks in each row and mark them as in Fig. 6(e). The different horizontal displacements are mapped to depth values using Eq. (11). This provides the depth map of one column of the stair target. By translating the scanning lens, we can compute the depth map of different columns of the scene and combine them into the 3D map visualized in Fig. 6(f). Fig. 6(g) provides a side view of the same target. In practice we also filter the reconstructed depth map to smooth it, as detailed in supplement.

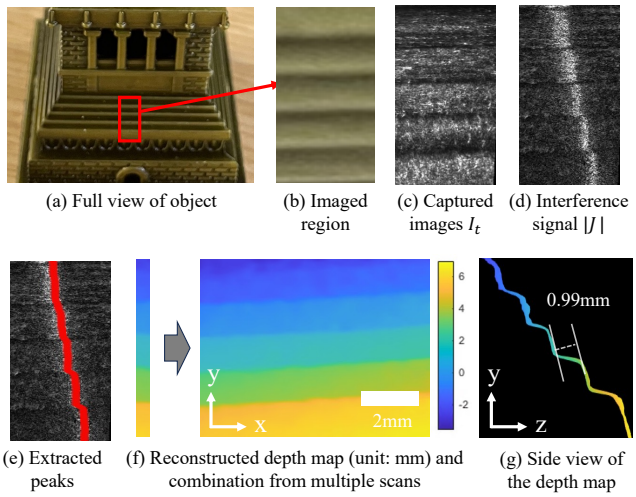


Figure 6. **Reconstruction steps.** (a) A zoom out image of the target, a metallic staircase. (b) The region we image in practice, with uniform appearance. (c) One input image I_t captured by our interferometer. (d) By combining 4 I_t images with different phase shifts ϕ_t we extract the interference component J. In each line there is a single bright region, the shift of the bright spot is the depth-dependent disparity shift. (e) We extract the peak of each row of J to compute depth. (f) Combining depth maps from multiple vertical scans provide the depth map of the full target. (g) A side view of the reconstructed target. The average measured distance between stair planes is around 0.99mm, while the ground truth is 0.9mm.

4. Resolution and range analysis

Our goal in this section is to understand the limits on resolution and range of the depths we can measure with our system. The supplementary file contains a detailed analysis leading to *analytic* formulas quantifying both for resolution and range, which is also carefully validated against numerical simulations. Due to space constraints, we only summarize here the main conclusions.

4.1. General interferometric considerations

Before diving into the specifics, we acknowledge general constraints on interferometer systems, as were derived

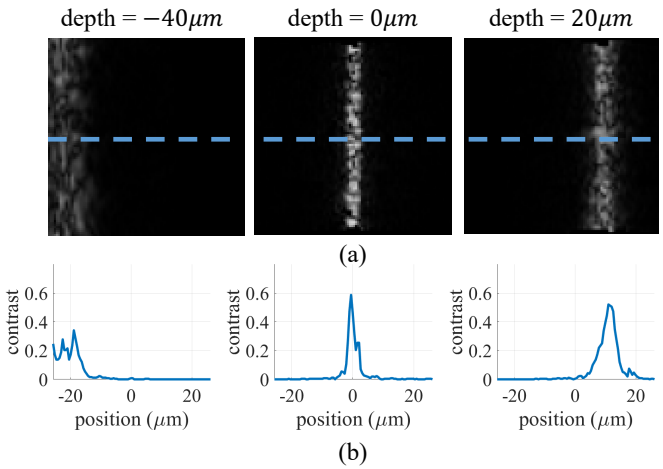


Figure 7. **Contrast in different depth planes.** We simulate the interference images J produced by planar targets at different depths. (a) Visualize the full image, and (b) visualizes a cross section. For planes father than the focal depth $z = 0$ we have wider defocus blur and the interference contrast reduces.

by [3]. These imaging conditions should be selected so that the interference term has a sufficient contrast to be detected and is not blurred out by the sensor. The interference contrast depends on three terms: the spatial coherence width of the illumination, the diffraction blur and the pixel pitch.

The coherence length Δc defines the range of spatial shifts over which the wavefront is still coherent with itself. As derived in [4, 11, 12], the spatial coherent length Δc is inversely proportional to the subtended angle to the light source. For example, sunlight has $\Delta c \approx 60\mu m$ as the sun is far away from the earth; on the other hand, for the indoor light source, Δc can be only a few microns wide.

We denote the pitch of the imaging sensor by Δx and the width of the diffraction blur by $\Delta\Phi$. $\Delta\Phi$ is inversely proportional to the aperture width.

The effect of these parameters on the contrast of the interference signal is carefully analyzed in [3], who conclude that to maximize interference contrast we should have

$$\Delta x < \Delta\Phi < \Delta c. \tag{13}$$

4.2. Range and resolution in our system

To gain intuition, consider the simulations in Fig. 7(a) where we simulate three planar targets at three different depths and visualize the interference images J they produce. These are indeed three vertical patterns with different displacements corresponding to their depths. Beyond the shift, when we move away from the focal plane of the camera, the interference pattern undergoes defocus blur which widens its extant and reduces it contrast.

Depth range. The range of depths we can measure is bounded, since the contrast of the interference patterns for far-away planes is too low to be reliably detected. We can analyze the contrast of the measured interference and arrive at the analytic bound summarized in the claim below. The bound depends also on the magnification M of the relay lens before the main imaging system (see part (c) of Fig. 5).

Claim 1 *We can measure objects in the depth range:*

$$|z| \leq \Omega_z, \quad \textit{with} \quad \Omega_z = \frac{1}{M}\frac{3\Delta\Phi\Delta c}{\lambda}. \tag{14}$$

The supplementary file provides the derivation underlying this claim. From this equation, we can see that, unsurprisingly, the depth range we can cover increases with $\Delta\Phi$. This corresponds to using a narrow aperture, where we can cover a larger depth of field with a smaller defocus blur. Beyond the poor light efficiency, there is a limit on our ability to reduce aperture size, since, as summarized in Eq. (13), aperture size is also bounded by coherence length $\Delta\Phi < \Delta c$.

Depth resolution. As derived in Eq. (12), the estimated depth is $\hat{z} = x_p/\beta$. 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, limiting the resolution of the estimated depth. This allows us to derive the following analytic formula for the achievable resolution (as before, we provide the derivation in the supplement).

Claim 2 *The depth resolution our system can resolve is*

$$\Delta z \approx \frac{1}{M}\frac{0.3\Delta c}{\beta}. \tag{15}$$

We note that the depth estimate in Eq. (12) is scaled by the tilt angle we image. Thus, it is expected that increasing the tilt angle β improves depth resolution. However, in practice, wide angles are more susceptible to optical aberrations and thus we cannot use arbitrarily large angles.

Number of depth planes. By dividing Eq. (14) and Eq. (22), we can conclude that the number of different depth planes we can resolve within the range is $\lambda^{-1}10\beta\Delta\Phi$. Since we are constrained to use $\Delta\Phi < \Delta c$ (Eq. (13)), we can detect at most

$$\frac{10\beta\Delta c}{\lambda} \tag{16}$$

different depth planes. To understand typical numbers, assuming e.g. $\beta = 0.57$ and $\Delta c = 4\mu m$, which corresponds to the coherence length of an indoor illumination, we can resolve ≈ 38 depth planes. On the other hand, sunlight has a much longer coherence $\Delta c \approx 60\mu m$, with which we can resolve ≈ 500 depth planes.

5. Hardware experiments

5.1. Prototype

We implement a hardware prototype, following the schematic of Fig. 5. Instead of translating the mirror, we make two arms in the interferometer orthogonally polarized, and use an LC cell to delay one arm. The supplementary provides more details as well as a components list.

For one-to-one magnification ($M = 1$) the size of the scene we can image is mostly set by the sensor size and it is about $8\times8\,\mathrm{mm}$ in our implementation. The depth range we can cover is also around $8\,\mathrm{mm}$. With a different magnification of the relay lens we can capture larger targets, equally scaling the size of the target area along all 3 axes.

In the following results, we used three different types of spatially incoherent illumination: 1) outdoor sunlight illumination, 2) indoor white light, and 3) a swept-angle laser illumination [9]. The first two sources are broadband and we added a bandpass filter, centered at $\lambda = 633\,\mathrm{nm}$ with FWHM of $5\,\mathrm{nm}$, to our system to make them monochromatic. The swept- angle illumination is not a natural one, but is included here for analysis purpose. Its advantage is that we can precise control the coherence length.

5.2. Range and resolution evaluation

We use our hardware prototype to evaluate contrast and resolution as in Sec. 4. We use a planar aluminum target and vary its depth using a motorized translation stage. We used the swept-angle illumination with controlled coherence $\Delta c = 108\,\mu\mathrm{m}$, diffraction blur kernel $\Delta\Phi = 8\,\mu\mathrm{m}$, and view angle $\beta \sim 0.03$. In Fig. 8(a), we show one interference image (at one depth). In Fig. 8(b) we plot the interference contrast as a function of depth. As predicted by our analysis in the supplementary file, the contrast reduces as we move away from the focal depth. In Fig. 8(c) we plot the mean peak position $\bar{x}_p$ as a function of the position of the target plane. The peak position, which is essentially a disparity shift, moves linearly with depth as expected in an orthographic imaging setup. The bars around the curve indicate the variance of the peak position and, as predicted in the supplementary file, this variance is larger when we move away from the focal plane. To estimate the variance and mean peak position, we image a planar target parallel to the camera, so all rows have the same depth. We detect the peak position in each row independently and then compute the mean and variance of these independent estimates.

5.3. 3D results with a swept-angle illumination

We work with spatial coherence $\Delta c = 54\,\mu\mathrm{m}$ to match the expected coherence length of the sunlight, diffraction blur kernel $\Delta\Phi = 8\,\mu\mathrm{m}$, and view angle $\beta \sim 0.05$. Since β here is small, the conclusions of Eq. (16) imply we can only resolve ≈ 6.3 depth planes. However, as described in supplement, for smooth targets we can average nearby points to reduce noise in the estimated x_p position and improve depth resolution. We also threshold the measured interference and display depth measurements only in image areas whose interference contrast is above some minimal value.

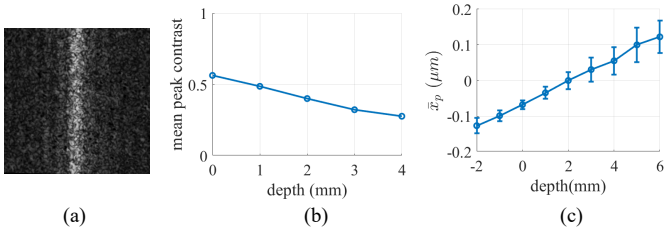
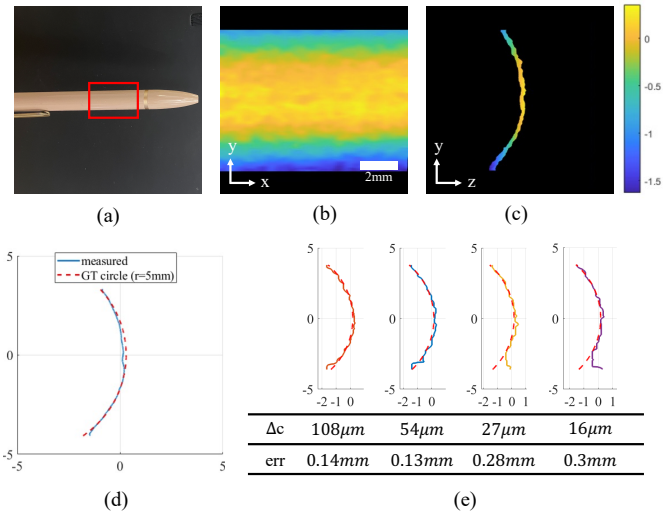


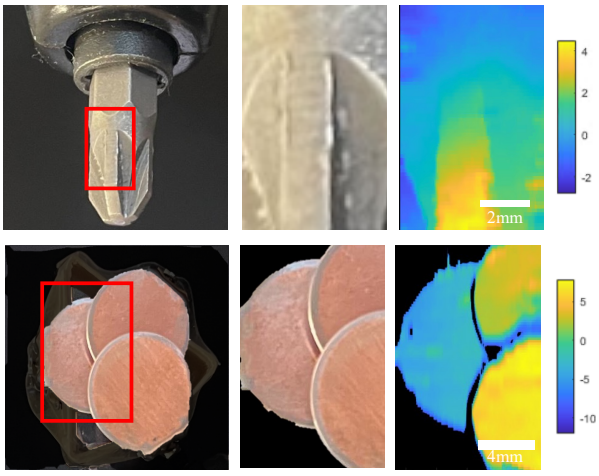
Figure 8. **Disparity and contrast as a function of depth.** We place a planar target at different depths and evaluate disparity and contrast. (a) Interference amplitude captured by our setup using a planar target. (b) The mean peak contrast as a function of the target plane depth. Contrast reduces when moving from the focal plane. (c) The mean peak position $\bar{x}_p$ w.r.t the target plane depth varies linearly. The variance of peak position, marked using bars around the main curve, also increases away from the focal plane.



Δc	108μm	54μm	27μm	16μm
err	0.14mm	0.13mm	0.28mm	0.3mm

Figure 9. **Reconstruction as a function of coherence length.** (a) A natural image of the target, a cylindrical pen, marking the reconstructed region. (b) Depth reconstruction. (c) A side view of the reconstruction showing the circular curve. (d) The measured depth map fitted with the expected circular shape. The average error is around 0.1 mm. (e) We repeat the experiment multiple times under different coherence lengths Δc and, as expected, the error increases with decreasing Δc.

In Fig. 9, we successfully reconstruct the cylindrical shape of a metal pen, which matches the expected circular shape at a high accuracy. We also repeat the experiments under different Δc. We can see that for smaller Δc, the error increase especially for the further points at the back of the pen. This matches the theory stating that the supported depth range decreases when coherence length decrease. In



(a) Reference full view of the target (b) Reference imaged region (c) Reconstructed depth map

Figure 10. **Reconstruction results under sunlight illumination.** Upper row: a screwdriver reconstructed using magnification $M = 1$. Lower row: planar circles reconstructed using $M = 0.5$.

Fig. 6, we also clearly reconstruct a stair target. We include additional reconstruction results in the supplement.

5.4. 3D results under natural illumination

Finally, we demonstrate our approach can reconstruct textureless targets under natural light. In Fig. 1, we use our setup to reconstruct the grind aluminum plane target. The target is illuminated by a lamp $1.25\,\mathrm{cm}$ wide and placed $50\,\mathrm{cm}$ away, which results in $\Delta c \sim 25\,\mu\mathrm{m}$. We can reconstruct the tilt plane shape of the target, despite the fact that no texture is visible to a standard intensity camera. In Fig. 10, we passively reconstruct targets under sunlight.

6. Limitations

The proposed approach uses interference and its measured contrast for passive shape reconstruction of textureless objects. We now discuss the dominant factors that reduce this contrast and hence degrade the depth estimation.

Coherence length. The interference contrast reduces when different coherent units of the wave are mixed in one measured pixel. For high contrast, the coherent length should be greater than the diffraction blur kernel width, which in turn is greater than the sensor pitch. For most cameras, the sensor pitch is at the order of a few micron. We can obtain a larger diffraction kernel by using a small aperture, at the cost of reduced light efficiency, and in practice our prototype uses an $f/8$ aperture. However, coherence length of the illuminant is scene specific. For sunlight, the coherence length is approximately of $60\,\mu\mathrm{m}$. Indoor lights can have wildly varying coherence length, depending on the spatial extents of the light sources. A standard light bulb, few meters away, has a coherence length that is approximately $16\,\mu\mathrm{m}$ in visible wavelengths [12]. More extended light sources can have very small coherence lengths, which will significantly reduce interference contrast. Magnification of the scene also changes coherence length. Imaging a large object would require using magnification ratios that are smaller than one; this results in a commensurate reduction in the coherence length.

Subsurface scattering. Our analysis in Sec. 4 ignores subsurface scattering. In practice, subsurface scattering "blurs" the wavefront, leaking one incoherent component into a neighboring one and reducing interference contrast. To minimize this problem, Most results in this paper used metallic subjects. The supplementary files has results with non-metallic objects, including sub-surface scatterers, showing the degradation in shape recovery.

Specular objects. The use of a shifted aperture results in the two arms of the proposed interferometer to observe a scene point from two distinct viewing directions. Objects that have view-dependent reflectance—such as shiny or specular objects—will result in the light passing through the two arms of the interferometer to have different levels of intensities. It is well-known in interferometry that the measured contrast is maximized when the two arms have similar light levels. This indicates that the performance of our approach will suffer for specular objects.

7. Conclusion

The results in this paper questions conventional wisdom in computer vision—*that, the shape of textureless objects cannot be passive reconstructed.* We rely on the observation that, even if no intensity differences are visible on the surface, the different incoherent components of the illumination can be isolated. This is made possible by using self interference between two viewpoints of a scene point, so as to measure its disparity. While our approach has limitations, primarily stemming from coherence length of the scene illumination, we hope it will spur renewed interest in self-interference and its possibilities.

Acknowledgment. We thank the support from the Israel Science Foundation (1947/20), the United States-Israel Binational Science Foundation (2008123, 2019758), the European Research Council (635537), Carnegie Mellon University Wei Shen and Xuehong Zhang Presidential Fellowship, the National Science Foundation (1730147, CAREER award 2238485). We also thank Alankar Kotwal and Ioannis Gkioulekas for their invaluable help in preparing the swept-angle light source and discussions.

References

[1] Oliver Cossairt, Nathan Matsuda, and Mohit Gupta. Digital refocusing with incoherent holography. In *ICCP*, 2014. 3, 1

[2] Jason Geng. Structured-light 3d surface imaging: a tutorial. *Advances in optics and photonics*, 3(2):128–160, 2011. 1

[3] Ioannis Gkioulekas, Anat Levin, Frédo Durand, and Todd Zickler. Micron-scale light transport decomposition using interferometry. *ACM TOG*, 34(4):1–14, 2015. 6

[4] Daniel Glasner, Todd Zickler, and Anat Levin. A reflectance display. *ACM TOG*, 33(4):1–12, 2014. 6

[5] Mohit Gupta, Ramesh Raskar, Achuta Kadambi, and Ayush Bhandari. Computational time-of-flight. http://web.media.mit.edu/~achoo/iccvtoftutorial. 1

[6] P Hariharan, BF Oreb, and Tomoaki Eiju. Digital phase-shifting interferometry: a simple error-compensating phase calculation algorithm. *Applied Optics*, 26(13):2504–2506, 1987. 2

[7] Berthold Klaus Paul Horn. Focusing. 1968. 1

[8] David Huang, Eric A Swanson, Charles P Lin, Joel S Schuman, William G Stinson, Warren Chang, Michael R Hee, Thomas Flotte, Kenton Gregory, Carmen A Puliafito, et al. Optical coherence tomography. *science*, 254(5035):1178–1181, 1991. 1

[9] Alankar Kotwal, Anat Levin, and Ioannis Gkioulekas. Interferometric transmission probing with coded mutual intensity. *ACM TOG*, 39(4):74–1, 2020. 4, 7, 6

[10] Alankar Kotwal, Anat Levin, and Ioannis Gkioulekas. Passive micron-scale time-of-flight with sunlight interferometry. In *CVPR*, 2023. 1

[11] Alankar Kotwal, Anat Levin, and Ioannis Gkioulekas. Swept-angle synthetic wavelength interferometry. In *CVPR*, 2023. 6

[12] Anat Levin, Daniel Glasner, Ying Xiong, Frédo Durand, William Freeman, Wojciech Matusik, and Todd Zickler. Fabricating brdfs at high spatial resolution using wave optics. *ACM TOG*, 32(4):1–14, 2013. 6, 8

[13] Martin D. Levine, Douglas A. O'Handley, and Gary M. Yagi. Computer determination of depth maps. *Computer Graphics and Image Processing*, 2(2):131–150, 1973. 1

[14] Bruce D Lucas and Takeo Kanade. An iterative image registration technique with an application to stereo vision. In *IJCAI*, 1981. 1

[15] Ramakant Nevatia. Depth measurement by motion stereo. *Computer Graphics and Image Processing*, 5(2):203–214, 1976. 1

[16] Alex Paul Pentland. A new sense for depth of field. *IEEE TPAMI*, (4):523–531, 1987. 1

[17] Joseph Rosen and Gary Brooker. Digital spatially incoherent fresnel holography. *Optics letters*, 32(8):912–914, 2007. 3, 1

[18] Joseph Rosen, A Vijayakumar, Manoj Kumar, Mani Ratnam Rai, Roy Kelner, Yuval Kashter, Angika Bulbul, and Saswata Mukherjee. Recent advances in self-interference incoherent digital holography. *Advances in Optics and Photonics*, 11 (1):1–66, 2019. 3

[19] Boxin Shi, Jinfa Yang, Jinwei Chen, Ruihua Zhang, and Rui Chen. Recent progress in shape from polarization. *Advances in Photometric 3D-Reconstruction*, pages 177–203, 2020. 1

[20] Hari Sundaram and Shree Nayar. Are textureless scenes recoverable? In *CVPR*, 1997. 1

[21] A Vijayakumar, Yuval Kashter, Roy Kelner, and Joseph Rosen. Coded aperture correlation holography–a new type of incoherent digital holograms. *Optics Express*, 24(11): 12430–12441, 2016. 3, 1

[22] Lawrence B Wolff and Terrance E Boult. Constraining object features using a polarization reflectance model. *Phys. Based Vis. Princ. Pract. Radiom*, 1:167, 1993. 1

[23] Robert J Woodham. Photometric method for determining surface orientation from multiple images. *Optical engineering*, 19(1):139–144, 1980. 1

[24] Ruo Zhang, Ping-Sing Tsai, James Edwin Cryer, and Mubarak Shah. Shape-from-shading: a survey. *IEEE TPAMI*, 21(8):690–706, 1999. 1

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.

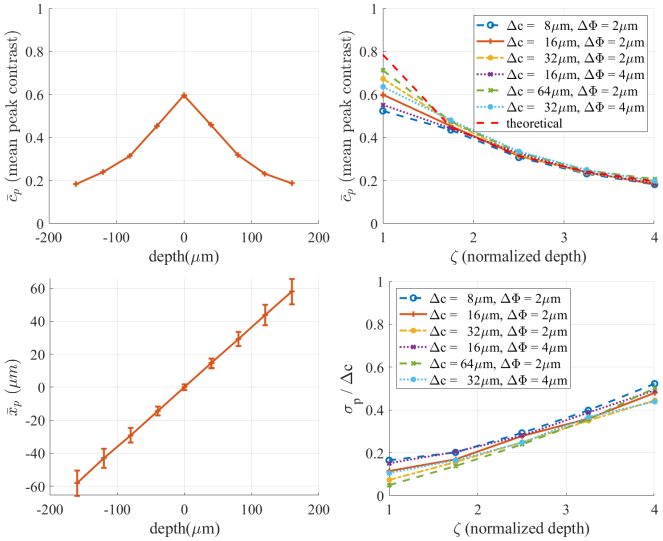


Figure 11. **Numerical analysis of range and resolution** (a) The contrast evaluated with the numerical simulation of Sec. 9.1, using planar targets at different depths. Contrast is highest when target is in focus (depth = 0), and decreases at larger distances. (b) Contrast for different Δc and $\Delta\Phi$ 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 $\frac{\lambda z}{\Delta\Phi}$. □

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

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 $\Delta z/M$.

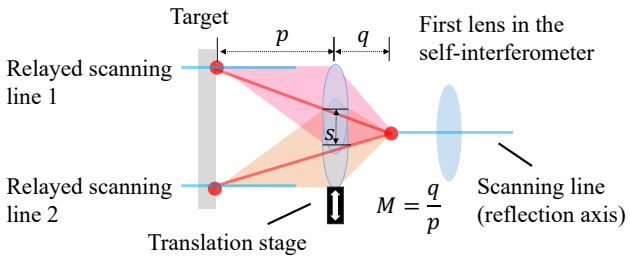


Figure 12. **Translating an additional lens to scan the scene.** Since our main setup can only scan a vertical slice of the scene, we add a relay lens in front of the main setup. By translating this lens we can scan different lines of the target, since a different strip of the scene is mapped to the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

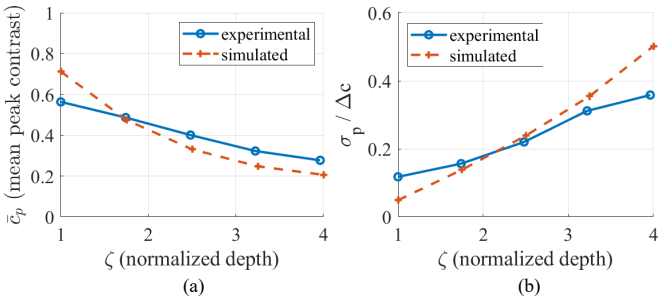
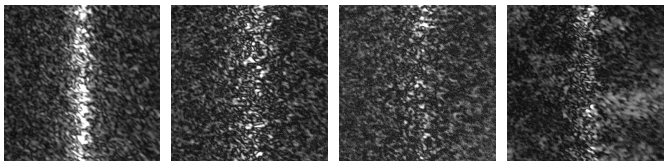


Figure 13. **Comparison between real and simulated results.** (a) Comparing contrast. (b) Comparing the standard deviation of peak position σ_p. While we do not know precisely all the parameters of the real system, the measured and simulated curves follow a similar behavior.

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.

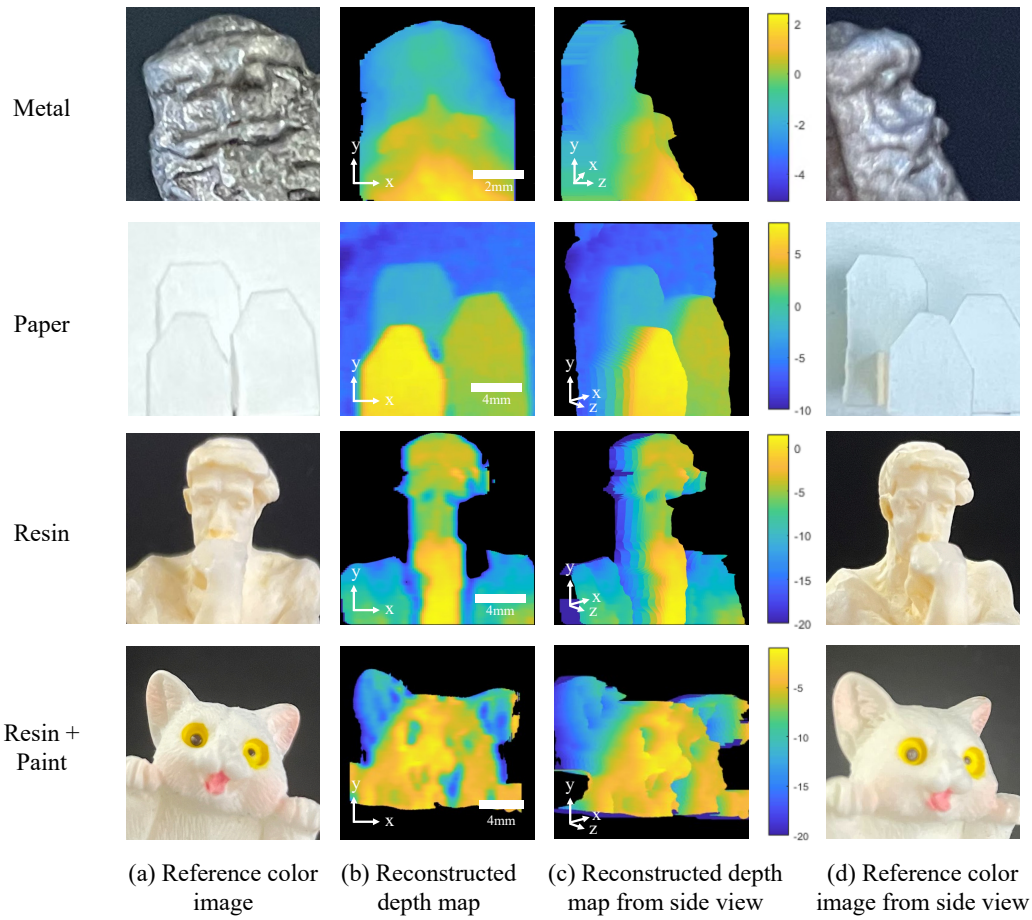


Figure 15. **Reconstruction results with swept-angle light source.** Targets in different rows are made of different materials. Targets with stronger subsurface scattering (paper, resin) have reduced depth resolution and can have artifacts when texture or material changes.

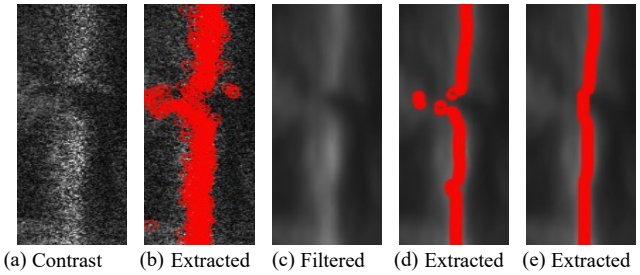


Figure 16. **Steps of the peak detection algorithm.** (a) The captured interference image, dominated by speckle noise. (b) Detecting the peak position in every row independently leads to noisy results. (c) Filtering the interference image (d) Peaks detected from the filtered images are smoother. (e)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

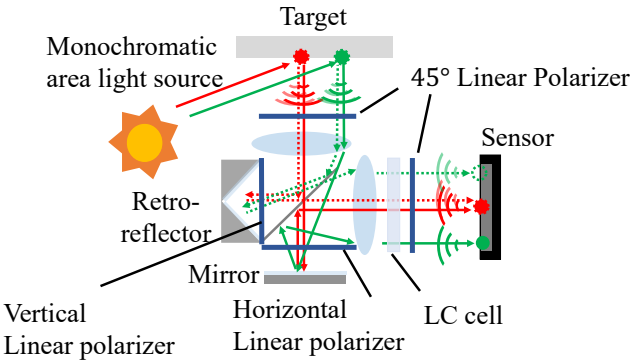
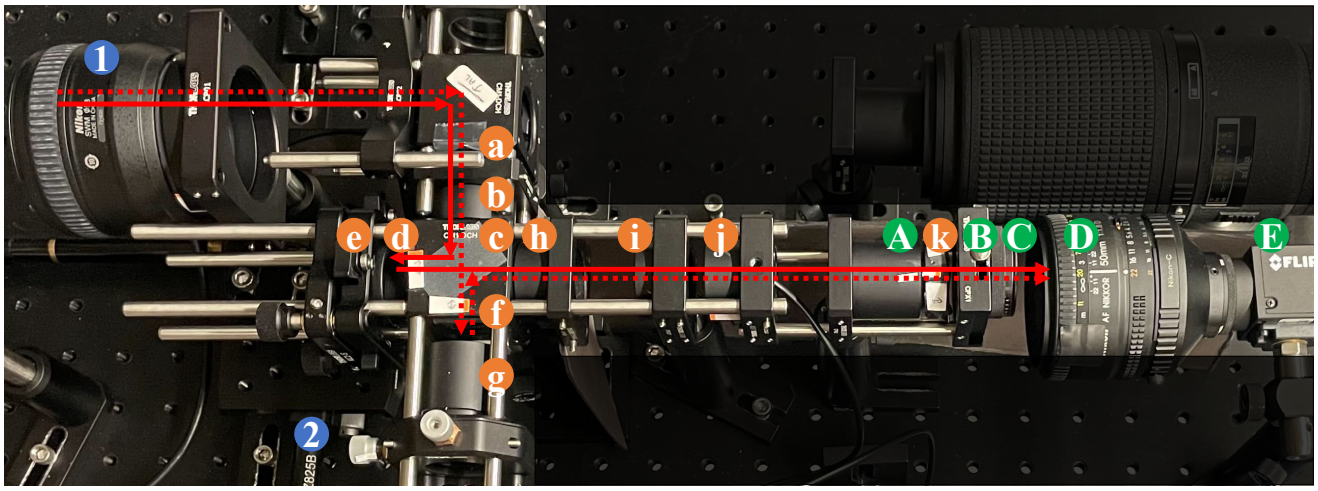


Figure 17. **Polarized interferometer.** Our interferometer is implemented using polarization rather than a translating mirror. The two arms are designed to have orthogonal polarization and used an LC cell to delay only the horizontally polarized waves.



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.