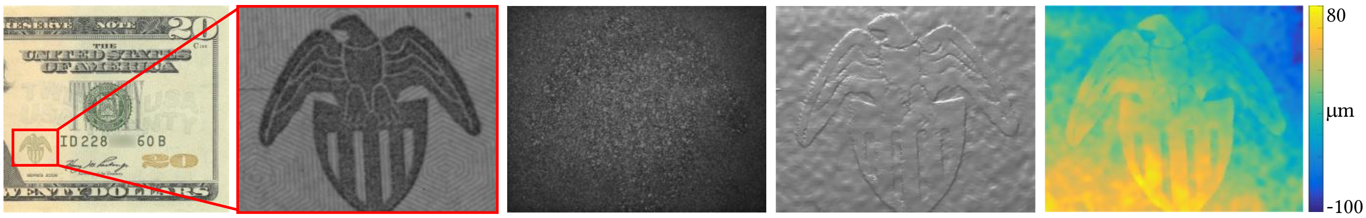
Swept-Angle Synthetic Wavelength Interferometry

ALANKAR KOTWAL, Carnegie Mellon University
ANAT LEVIN, Technion
IOANNIS GKIOULEKAS, Carnegie Mellon University



dollar bill | ambient light image | captured raw image | recovered depth surface | recovered depth map

Fig. 1. **Reconstructing the eagle embossed on a United States twenty-dollar bill**. The features on the eagle are raised 20 µm off the surface of the bill. From left to right: location of the eagle on the bill, an ambient light image of the eagle, a raw frame captured by our method, the depth recovered by our method rendered as a surface, and plotted as a depth map. The state-of-the-art technique, optical coherence tomography, requires twenty times as many measurements as our method does to capture a depth map of comparable quality.

Phase-shifting interferometry (PSI), optical coherence tomography (OCT) and amplitude-modulated time-of-flight (AMCW-ToF) systems are some of the most reliable and commercially successful technologies for measuring shapes of objects at nanometer, micrometer and millimeter axial resolutions respectively. Other than axial resolution, existing implementations of these present trade-offs against each other: Whereas AMCW-ToF systems promise fast, full-field acquisition and tunability of unambiguous depth range, spatial scanning-based swept-source OCT and PSI promise high spatial resolution and resilience to global illumination caused by multibounce light transport.

We present an interferometric method that combines the best of these worlds, presenting fast, full-field micrometer-resolution and range-tunable depth acquisition robust to global illumination. We construct a measurement model analogous to PSI and AMCW-ToF using light consisting of two optical wavelengths, generating in the interferometric measurements a per-pixel phase that encodes its depth. This phase can be measured over the entire camera at once in as little as 16 measurements, yielding fast acquisition. The separation between these optical wavelength tunes the range of unambiguous depth measurement. In addition to these, we construct spatially incoherent illumination that provides these interferometric measurements pixel-level robustness to global illumination. We term this method swept-angle synthetic wavelength interferometry.

We demonstrate the capabilities of our method by capturing high spatial- and axial-resolution depth for a variety of scenes under challenging light transport effects such as interreflections, subsurface scattering and specularities. We validate our results by comparing them against those captured with optical coherence tomography, the standard for micron-scale depth acquisition. In addition, we quantify the numerical accuracy of our estimates and demonstrate the robustness of our method to ambient light.

CCS Concepts: • **Computing methodologies** → **3D imaging**; **Computational photography**.

Additional Key Words and Phrases: 3D sensing, interferometry, probing

1 INTRODUCTION

Depth sensing is a cornerstone of computer vision research due to its applications in biomedical imaging, industrial fabrication, robotic perception, autonomous driving and human-computer interaction. As a result, there is a huge body of work upon acquiring the 3D shapes of objects in computer vision and computational imaging. In particular, micrometer-resolution depth sensing is important in biomedical imaging because biological features are often micron-scale, industrial fabrication to manufacture critical parts within microns of their specifications, and robotics to handle fine objects.

Among depth sensing methods, *active illumination* methods have gained favor in recent times. Their control upon scene illumination guarantees resilience against most scene texture and reflectance properties and ambient light conditions. Active illumination techniques have been adopted in autonomous navigation (lidar), gaming (structured light stereo and continuous-wave time-of-flight in Kinect), face recognition (structured light in iPhone Face ID) and biomedical imaging and industrial fabrication (interferometry with optical coherence tomography). All these modalities share largely the same operating principle, comparing the response of the scene to the illumination with the illumination itself.

Some of these active illumination techniques provide direct estimates of per-pixel depth without any computational overhead. Among these, amplitude-modulated continuous-wave time-of-flight (AMCW-ToF) methods perform temporal photo-electric modulation of the light waveform projected upon the scene and calculate the shift of the incoming waveform to estimate pixel depth. However, due to constraints on the electronics performing modulation, the axial resolution of these methods is restricted in practice to millimeters. In addition, the unambiguous depth range of AMCW-ToF measurements is restricted to the wavelength of the modulation. On the other end of the spectrum, phase-shifting interferometry (PSI) techniques use the waveform of monochromatic light and the phenomenon of interference to construct a similar measurement model, measuring depths at nanometer resolutions. This comes at the price of tiny unambiguous depth ranges, of the order of a micron.

Synthetic wavelength interferometry (SWI) methods bridge the gap between micron PSI wavelengths and meter AMCW-ToF wavelengths by synthesizing modulation with an intermediate wavelength. They use light consisting of two narrowly-separated optical wavelengths and the phenomenon of interference, and the synthesized wavelength depends upon the separation between the optical wavelengths. They provide a clear way of trading off unambiguous depth range with depth resolution. In particular, they provide us a way to reach our target of micron-resolution depth sensing without restricting unambiguous depth range to micrometers.

The current standard for micron-scale depth sensing is optical coherence tomography (OCT). OCT uses broadband illumination to measure the temporally-resolved response of the scene and reconstructs depth as the position of the peak of this response. However, OCT presents a huge time cost compared to SWI: The number of measurements OCT takes is proportional to the depth range and inversely proportional to the required axial resolution. In contrast, SWI can achieve the same in as little as 16 measurements

All these measurement models assume single-bounce light transport in the scene. However, in reality, light takes many different paths in the scene, rendering the measurement model inaccurate. This is especially problematic in measuring high-resolution depth: Therefore, traditional implementations of OCT and PSI have been based in fiber optics, with single points on the scene being scanned and measured sequentially. While this provides robustness to global illumination, it slows down acquisition and reduces spatial resolution significantly compared to full-field imaging. We adopt one of the results by Gkioulekas et al. [2015] to reject global illumination in the interference measured even in full-field imaging.

Our main contribution in this work is to show that a combination of synthetic wavelength interferometry with the Fourier-domain redistributive projector of [Kotwal et al. 2020] yields fast, full-field, range-tunable depth sensing robust to multibounce light transport. We call this combination swept-angle synthetic wavelength interferometry. We begin with some background on PSI and extend it to SWI. Then, we explain the benefits and disadvantages of fiber-based and full-field implementations of these in the presence of global illumination. Then, we state and prove in the supplement the invariance of swept-angle synthetic wavelength interferometry to global illumination even in full-field imaging. We present an acquisition and reconstruction pipeline for our method, and finally demonstrate the performance, micron spatial and axial resolution, and robustness of our method in the face of complications presented by multibounce light transport.

2 RELATED WORK

Below we review related approaches in the context of depth acquisition and the elimination of global illumination effects.

2.1 Depth acquisition

Passive depth sensing. Passive methods in depth sensing rely on the appearance of the scene under external, ambient light. These methods exploit image cues such as disparity in multiview geometry [Barnard and Thompson 1980; Hartley and Zisserman 2004; Nalpantidis et al. 2008], camera defocus [Grossmann 1987; Hazirbas et al. 2018; Subbarao and Surya 1994], or shading [Han et al. 2013; Horn 1970], to name a few examples. All these methods permit inexpensive implementation involving no human intervention in the scene and using ordinary cameras. However, they heavily depend upon the presence of texture in the scene, and their resolution is limited to the resolution at which such features can be detected.

Structured light. One method to circumvent the lack of texture is to actively project custom-generated light on the scene [Chen et al. 2008; Gupta et al. 2011; O'Toole et al. 2016; Scharstein and Szeliski 2003]. However the spatial resolution of structured light is restricted by the resolution of the projected patterns, and usually does not achieve sub-millimeter axial and spatial resolution.

Time-of-flight. Rather than relying on spatially-varying cues, time-of-flight sensors recover depth encoded as the delay between emitted and received waveforms. *Pulsed* time-of-flight sensors send out pulses of laser light into the scene and detect their return via ultrafast photodiodes [Kirmani et al. 2009], Geiger-mode avalanche photodiodes [Aull 2005; Kirmani et al. 2014], single-photon avalanche diodes [Gariepy et al. 2015; Gupta et al. 2019a,b; Heide et al. 2018; Lindell et al. 2018; O'Toole et al. 2017] and streak sensors [Velten et al. 2012]. The axial resolution of these methods is limited by the temporal resolution of the sensors, usually leading to depth accuracy in the order of centimeters, or millimeters with the aid of computational reconstruction [Heide et al. 2018; Velten et al. 2012]. In addition, the low spatial resolution of available arrays of such ultrafast sensors [Niclass et al. 2005; Rochas et al. 2003; Villa et al. 2014] denies us the required micron-scale spatial resolution.

Related more closely to our technique are amplitude-modulated continuous-wave time-of-flight (AMCW-ToF) sensors. These sensors operate by sending out light modulated by a periodic function of time, and estimating the shift of the returning waveform with respect to the emitted waveform [Ferriere et al. 2008; Flores et al. 2014; Gupta et al. 2018; Gutierrez-Barragan et al. 2019; Lange and Seitz 2001; Payne et al. 2011; Piatti et al. 2013]. With the advent of photonic mixer devices [Heide et al. 2013; Schwarte et al. 1997] and per-pixel photo-demodulators [Lange and Seitz 2001; Lange et al. 2000], these sensors have become the modality of choice for measuring millimeter-resolution depth and have been applied in popular products such as the Kinect for Xbox One. The range tunability and axial resolution of these sensors is restricted by the maximum modulation and demodulation frequencies allowed by these technologies, corresponding to a millimeter-scale resolution. In addition, per-pixel demodulation-based sensors have low fill factors and large pixel sizes [Piatti et al. 2013], denying us micron-scale spatial resolution.

Optical interferometry. Interferometry is a classic wave-optics technique that measures the correlation, or *interference*, between two or more light beams that have traveled different along paths [Hariharan 2003]. Since interferometric effects are highly sensitive to wavelength-scale features, they are ideal for imaging at micron resolutions. *Phase-shifting interferometry* (PSI), which uses single-frequency light to perform nanometer-scale depth estimation [de Groot 2011; Johnson et al. 2001], can be thought of as a special case of AMCW-ToF, with the modulation frequency being the terahertz frequency of oscillation of the light waveform instead of photo-electric modulation. Interferometric methods encode temporal delay information in the difference in optical path length between the two paths, eliminating the need for demodulation electronics capable of operating at terahertz frequencies and making it possible to perform interferometric sensing with ordinary photodetectors and cameras. However, the unambiguous depth range of PSI is restricted to the wavelength of the light used, typically around a micron.

Heterodyne interferometry methods bridge the gap between the megahertz-scale modulation frequencies of AM-CW ToF systems and the terahertz-scale frequencies of interferometry by *synthesizing* modulation at an intermediate frequency [Cheng and Wyant 1984, 1985; de Groot and McGarvey 1992; Fercher et al. 1985; Li et al. 2018, 2017]. They give us the flexibility to trade off unambiguous depth range with resolution by tuning this synthetic wavelength. We build upon heterodyne interferometry, adding onto it the ability to reject an unwanted component of scene appearance called global illumination, which we will take a look at next.

Finally, *optical coherence tomography* (OCT) [Gkioulekas et al. 2015; Huang et al. 1991; Kotwal et al. 2020] is another interferometric technique that uses broadband illumination for micron-scale depth acquisition. This gives OCT the ability to decouple range and resolution, allowing unambiguous imaging arbitrary depth ranges at a user-picked resolution, albeit at a huge time cost. OCT can be implemented with a setup identical to the one we use in this paper: Hence, we use the same hardware to capture both OCT and swept-angle synthetic wavelength interferometry data, obtaining ground truth depth to compare against.

2.2 Mitigating global illumination

Global illumination. In their primitive forms, both time-of-flight and interferometric depth sensing techniques assume the absence of multipath light transport in the scene. Multipath light transport results in *global illumination*, that confounds information about the true time-of-flight in the acquired measurements, and therefore needs to be eliminated. There has been progress in *computationally* undoing global illumination in continuous-wave time-of-flight systems based on modeling multi-bounce light transport [Fuchs 2010; Jimeneza et al. 2014; Naik et al. 2015], sparse reconstruction [Freedman et al. 2014; Kadambi et al. 2013], multi-wavelength approaches [Bhandari et al. 2014] and neural approaches [Marco et al. 2017].

A second class of methods for *optically* removing global illumination is based on *probing light transport* [O'Toole et al. 2012]. Optical computation is superior to the computational methods mentioned above because it happens before photo-conversion by the sensor, eliminating the detrimental effect of sensor noise artifacts like read noise and quantization on the computational algorithm. In addition, it doesn't require explicitly modeling light propagation as some of the computational methods approximate. Light transport probing methods modulate, in addition to the temporal profile in time-of-flight, the spatial properties of the illumination and camera. The different kinds of spatial illumination and acquisition conditions used include epipolar imaging [Achar et al. 2017], high-spatial-frequency illumination [Nayar et al. 2006; Reddy et al. 2012] and swept-temporal-frequency illumination [Gupta et al. 2015; O'Toole et al. 2014a]. However, as in the structured light techniques mentioned before, the spatial resolution of these techniques is restricted by the maximum possible spatial frequency of projected illumination: The sub-resolution part of global illumination not removed reduces both the axial and spatial resolutions of the estimated depth.

The majority of interferometric depth sensing techniques rejects global illumination use *impulse illumination and imaging*. By illuminating and imaging one point at a time, global illumination arising from other parts of the scene is prevented. This is implemented using either collimated light from optical fibers [Huang et al. 1991; Li et al. 2018, 2017] or spatial focusing optics [Cheng and Wyant 1984, 1985; de Groot 2011; Hariharan 2003; Johnson et al. 2001; Meiners-Hagen et al. 2009]. The main disadvantage of this approach is the need to sequentially scan the impulse projection over the 2D scene. In contrast, we adapt here the transmission probing approach introduced by Gkioulekas et al. [2015] and Kotwal et al. [2020], which allow full-field imaging while rejecting global illumination.

3 BACKGROUND

We give here some background on interferometry in general, and synthetic wavelength interferometry in particular. We follow the analysis of Kotwal et al. [2020] and Gkioulekas et al. [2015].

The Michelson interferometer. Our optical setup is based on the classical Michelson interferometer. Figure 2 shows a schematic of this setup. The interferometer uses a beam splitter to divide collimated input illumination into two beams: one propagates toward the *scene arm* containing the scene of interest, and another propagates toward the *reference arm*, typically a planar mirror mounted on a translation stage that can vary the mirror's distance from the beam splitter. After reflection from both arms, the two light beams recombine at the beam splitter and propagate toward the sensor. The illumination input to the interferometer is typically created by placing a monochromatic point source, such as a single-frequency laser diode, in the focal plane of a collimating lens.

We denote by $\boldsymbol{u_s}(x)$ and $\boldsymbol{u_r}(x)$ the fields at sensor pixel x due to the scene and reference respectively. The sensor measures an image equal to the intensity of the superposition of the two fields,

$$I(x) = \|\boldsymbol{u_s}(x) + \boldsymbol{u_r}(x)\|^2 \tag{1}$$

$$= \|\boldsymbol{u_s}(x)\|^2 + \|\boldsymbol{u_r}(x)\|^2 + 2\,\mathrm{Re}\left\{\boldsymbol{u_s}(x)\,\boldsymbol{u_r}(x)^*\right\}. \tag{2}$$

The first two terms in Equation (2) are the intensities the sensor would measure if it were observing each of the two arms separately. The third term, which we term *interference*, is the real part of complex correlation $\boldsymbol{C}$ between the scene and reference fields,

$$\boldsymbol{C}(x) \equiv \boldsymbol{u_s}(x)\;\boldsymbol{u_r}(x)^*. \tag{3}$$

Interferometric methods focus on measuring and analyzing the interference [Abramson 1983; Creath 1985; Gkioulekas et al. 2015; Huang et al. 1991; Kotwal et al. 2020; Li et al. 2018, 2017; Maeda et al. 2018]. The interference can be isolated from Equation (2) in various ways, and we elaborate on one strategy in Section 4.

Phase-shifting interferometry. We now assume that the input illumination is a collimated monochromatic beam (for example, created by placing the output of a single-frequency laser at the focal plane of an aberration-corrected lens, as in Figure 2(a)). Then, we can model the input illumination as a plane wave traveling along the optical axis z of the interferometer,

$$\boldsymbol{u_i}(x, z) = \exp(-i\kappa z), \tag{4}$$

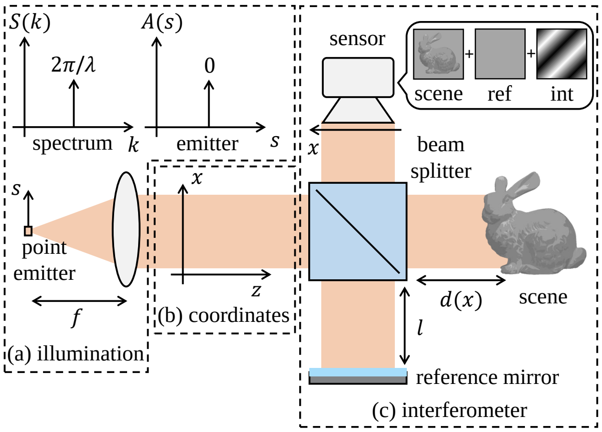


Fig. 2. **The Michelson interferometer and its input illumination**. **(a)** Illumination in a Michelson interferometer is created by placing a monochromatic point source, such as a single-frequency laser diode, emitting at wavelength λ in the focal plane of a collimating lens. Coordinates in the focal plane are labeled as s. The emitter then has a spatial amplitude profile $A(s) = \delta(s)$, and a spectrum $S(k') = \delta(k' - k)$, where $k = 2\pi/\lambda$. **(b)** Coordinates along the illumination injected in the interferometer: x labels the coordinates along the lens and sensor planes, while z labels distances along the optical axis. **(c)** A beam splitter divides the input illumination into two 'arms': One traveling to the scene, and one to the reference. Upon reflection from the two arms, light from both arms is superimposed on the camera. The camera measures a sum of three components: The individual intensities of the arms, and an 'interference' component. We're interested in the interference component.

where λ and $\kappa \equiv 2\pi/\lambda$ are the illumination *wavelength* and *wavenumber*, respectively. The correlation $\boldsymbol{C}$ of Equation (3) encodes information about the distance of the target scene from the beamsplitter. To understand this, we can consider the form of the scene and reference fields for illumination of the form of Equation (4). If the distance between the reference mirror and the beam splitter is l, the reference field is simply the input field shifted in phase:

$$\boldsymbol{u_r}(x) = \exp(-2i\kappa l). \tag{5}$$

Similarly, if the distance between the sensor pixel x images is $d(x)$, then the scene field equals

$$\boldsymbol{u_s}(x) = \exp(-2i\kappa d(x)). \tag{6}$$

Combining Equations (5)-(6) with Equation (3), the correlation is

$$\boldsymbol{C}(x, l) = \exp(-2i\kappa(d(x) - l)). \tag{7}$$

Therefore, the correlation encodes the $d(x)$ of the scene point in its phase $\phi(d(x)) \equiv 2\kappa(d(x) - l)$. Phase-shifting interferometry estimates this phase by using the translation stage to *shift* the reference mirror location by sub-wavelength amounts. Most commonly, phase-shifting interferometry uses measurements $\mathcal{I}(x, l_m)$ at four locations $l_m = l + m\lambda/8$, $m \in \{0, 1, 2, 3\}$ of the reference mirror. Up to a constant A in Equation 2, these measurements are the real part of the correlation: $\mathcal{I}(x, l_m) = A + 2\,\mathrm{Re}\{\boldsymbol{C}(x, l_m)\}$. These measurements $\mathcal{I}(x, l_m)$ can therefore estimate the phase of the sinusoid using the *four-bucket phase retrieval algorithm* [Bruning et al. 1974]:

$$\tan(2\kappa(d(x) - l)) = \frac{\mathcal{I}(x, l_3) - \mathcal{I}(x, l_1)}{\mathcal{I}(x, l_0) - \mathcal{I}(\mathcal{I}, l_2)}, \tag{8}$$

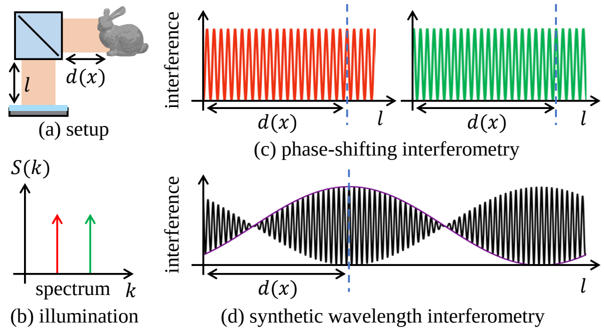


Fig. 3. **A visual illustration of synthetic wavelength interferometry**. **(a)** We consider a point x on the scene placed at a distance $d(x)$ from the beam splitter and the corresponding reference mirror point placed at a distance l. **(b)** The illumination injected into the interferometer consists of emission at two distinct, but narrowly-separated optical wavelengths, represented as the red and green peaks. **(c)** As a function of the distance l to the reference mirror, each peak contributes to the interference component a sinusoid with a period equal to its wavelength. Both of these sinusoids achieve one of their maxima at $l = d(x)$: Estimating the location of the maxima is the basis for phase-shifting interferometry using each individual wavelength. However, the depth range of phase-shifting interferometry is restricted to half the wavelength used. **(d)** The sum of these two can be expressed as a product of two waveforms: a sinusoid at one of the optical wavelengths drawn in black, and another at a *synthetic wavelength* peaked at $l = d(x)$ drawn in purple. Estimating the location of this peak is the basis for synthetic wavelength interferometry. This synthetic wavelength is typically orders of magnitude higher than the optical wavelengths. Synthetic wavelength interferometry, therefore, extends the depth range of phase-shifting interferometry while still achieving micron-scale resolution.

which gives for depth

$$d(x) = \frac{1}{2\kappa}\arctan\left[\frac{\mathcal{I}(x, l_3) - \mathcal{I}(x, l_1)}{\mathcal{I}(x, l_0) - \mathcal{I}(x, l_2)}\right] + l + \frac{n\lambda}{2}, \tag{9}$$

for any integer n. That is, one can measure the residual depth at intervals of $[0, \lambda/2]$, but cannot disambiguate between depths differing by an integer multiple of $\lambda/2$. For this reason, phase-shifting interferometry is limited to applications requiring sub-wavelength depth ranges, such as wavefront measurement [Koliopoulos 1981], optical surface characterization [Wyant et al. 1984], and optical instrument testing [Bruning et al. 1974].

Synthetic wavelength interferometry. To extend the unambiguous depth range beyond the optical wavelength λ, synthetic wavelength interferometry techniques use illumination at two distinct, but narrowly-separated, wavelengths that are incoherent with each other. We denote their wavenumbers as κ and $(1 + \epsilon)\kappa$, corresponding to wavelengths λ and $\lambda/1+\epsilon$. As the two wavelengths are incoherent, the correlation $\boldsymbol{C}(x, l)$ in the presence of both is the sum of the per-wavelength correlations from Equation (7):

$$\boldsymbol{C}(x, l) = \exp(-2i\kappa(d(x) - l)) + \exp(-2i\kappa(1 + \epsilon)(d(x) - l)) \tag{10}$$

$$= \exp(-2i\kappa(d(x) - l))\left[1 + \exp(-2i\kappa\epsilon(d(x) - l))\right]. \tag{11}$$

Since intensity measurements obtained by the camera give us the real part of $C(x, l)$, we will deal with its real part:

$$\mathrm{Re}\left(C(x,l)\right) = \cos(2\kappa(d(x)-l)) + \cos(2\kappa(1+\epsilon)(d(x)-l)) \quad (12)$$

$$= 2\sin(\kappa(2+\epsilon)(d(x)-l))\sin(\kappa\epsilon(d(x)-l)) \quad (13)$$

$$\approx 2\sin(2\kappa(d(x)-l))\sin(\kappa\epsilon(d(x)-l)), \quad (14)$$

where the approximation is accurate when $\epsilon \ll 1$. The correlation measured by our camera as a function of $d(x) - l$ is then a *carrier sinusoid* at wavelength $\lambda/2$, modulated in amplitude by another sinusoid with a *synthetic wavelength* $\lambda_s \equiv \lambda/\epsilon$. We term this sinusoid the *envelope* $\mathcal{E}$ of the correlation:

$$\mathcal{E}(x,l) \equiv \sin(\kappa\epsilon(d(x)-l)) \quad (15)$$

Figure 3 visualizes the correlation $\mathrm{Re}\{C(x,l)\}$, and envelope $\mathcal{E}(x,l)$ functions. We can estimate the phase of this sinusoid by capturing envelope measurements at four locations $l_m = l + m\lambda_s/4$, $m \in \{0,1,2,3\}$, and estimate from it the depth $d(x)$ using the same four-bucket algorithm expression as in Equation (9):

$$d(x) = \frac{1}{\kappa_s}\arctan\left[\frac{\mathcal{E}(x,l_3)-\mathcal{E}(x,l_1)}{\mathcal{E}(x,l_0)-\mathcal{E}(x,l_2)}\right] + l + n\lambda_s. \quad (16)$$

Full-field and scanning interferometry. There are broadly two types of Michelson inteferometer setups that can be used to implement both phase-shifting and synthetic wavelength interferometry: (a) a full-field interferometer; and (b) a scanning interferometer. We discuss the advantages and disadvantages of the two types, which will motivate our proposed swept-angle interferometer setup.

Full-field interferometers create a large beam to illuminate the entire area of interest in the scene arm, and corresponding area in the reference arm, all at once. Additionally, they use a two-dimensional sensor to measure the superposition of the reflected fields. Figure 2 shows a typical implementation of a full-field interferometer using *free-space optics* (lenses for beam generation and sensor focusing; beam splitter for separating and recombining fields). This interferometer can be modified easily to implement full-field synthetic wavelength interferometry by combining the outputs of two fiber-coupled single-frequency lasers and placing the fiber tip in the focal plane of the collimating lens. Such an interferometer is shown in Figure 5(a). We will assume for the rest of this section that our illumination consists of this two-wavelength spectrum.

Full-field interferometers enable fast measurements of correlation for all scene points at once, and at spatial resolutions as high as the pixel pitch of the sensor. However, these interferometers are susceptible to *indirect illumination* effects. To understand this, we consider the example scene of Figure 4, inspired from O'Toole et al. [2012]. By expressing the (per-wavelength) scene field $\boldsymbol{u_s}(x)$ as in Equation (6) and the envelope $\mathcal{E}(x,l)$ as in Equation (15), we are implicitly assuming that the only light path that contributes to the field is the *direct* light path (green path in Figure 4(a)): It originates at point x on the illumination places, reflects on the scene surface exactly once, and ends at the corresponding point x on the sensor plane. Such a path has a length of $2d(x)$, resulting in the corresponding phase delay in Equations (6), (15) for the scene field and envelope, respectively (green sinusoid in Figure 4(b)).

In practice, the scene field will include contributions from *indirect* light paths, marked with dashed lines in the figure. All these paths have path length different from $2d(x)$, and contribute to the

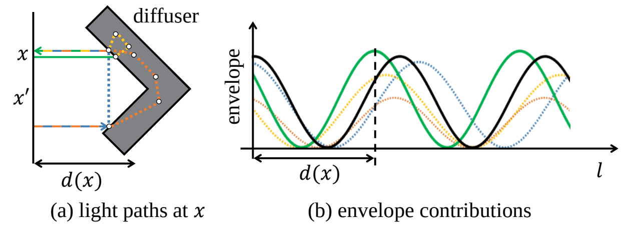


Fig. 4. **Full-field interferometry and global illumination**. **(a)** A typical scene contains two kinds of light paths: the direct path in green, and the variously-colored dashed global illumination paths. **(b)** The direct path has a length of exactly by the distance to the scene point, contributing the green sinusoid with the correct phase. The other paths have the wrong length, therefore contribute sinusoids with the wrong phase. The total synthetic wavelength sinusoid, marked in black, is the sum of all such sinusoids. It has the wrong phase, thus yielding an erroneous depth estimate.

envelope terms of different phase (differently-colored dashed sinusoids in Figure 4(b)). The camera measures an envelope that is the *superposition* of these sinusoids. This is another sinusoid with phase $2d' \neq 2d(x)$. Consequently, using the four-bucket algorithm of Equation (16) to estimate depth produces inaccurate results.

Scanning interferometers use a narrow or focused beam to illuminate only one point in the scene arm, and corresponding point in the reference arm, at any given time. Additionally, they use a single-pixel sensor, focused at the same scene and reference arm points, to measure the superposition of the reflected fields. To capture measurements of correlation of the entire scene, steering optics scan the focus point across the region of interest in the scene and reference arms. Figure 5(b) shows a typical implementation of a scanning interferometer using *fiber optics* (fiber collimators for beam generation and coupling, fiber splitters and circulators for separating and recombining fields), and beam steering optics (e.g., a MEMS mirror).

Scanning interferometers are very effective at mitigating the effects of indirect illumination because, at any given time, they only illuminate and image one point x in the scene, they eliminate contributions from indirect paths starting at different points x' and ending at x (blue and orange paths, respectively, in Figure 4(a)). Among the remaining paths that contribute to the scene field, the contribution of the direct path (green path in Figure 4(a)) dominates those of indirect paths that start and end at the same point (yellow path in Figure 4(a)), as explained by O'Toole et al. [2014b; 2012]. Measuring only these paths is equivalent to *diagonal probing* measurements in co-axial configurations, as stated in O'Toole et al. [2012].

Unfortunately, this robustness to indirect illumination comes at the cost of having to use beam steering to scan the entire scene. This creates several problems for applications where it is necessary to measure depth at micrometer-scale axial *and* spatial resolutions. In particular, realizing such high spatial resolution requires: (i) a laser beam a few micrometers wide; (ii) a MEMS mirror capable of scanning at high-enough angular resolution to translate the laser beam a few microns on the scene surface; and (iii) acquisition time long enough to scan a megapixel-size grid on the scene. These requirements are generally challenging, as we elaborate in the supplement.

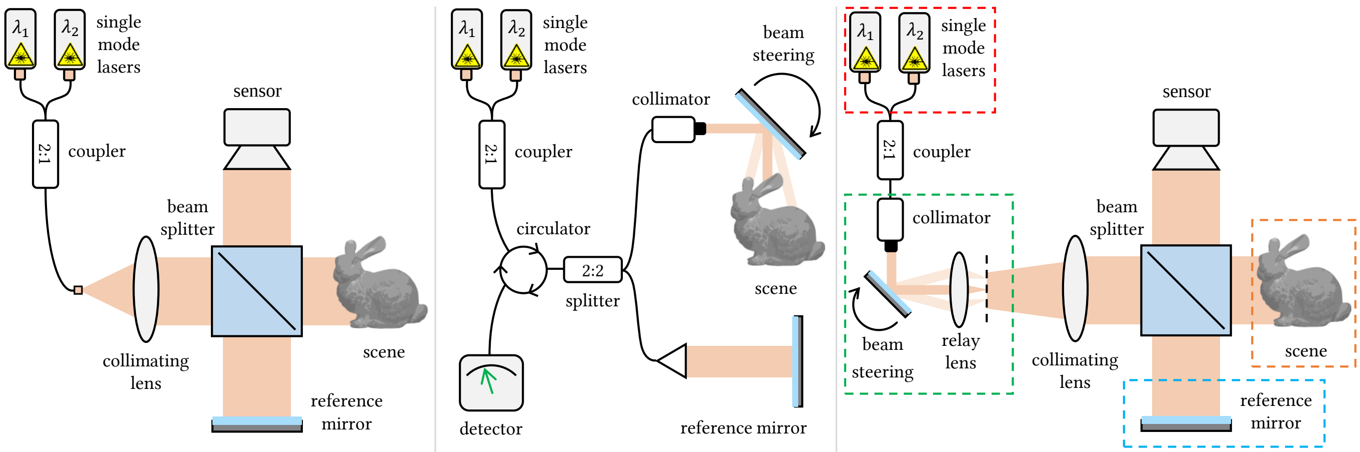


Fig. 5. **Schematics of hardware setups that implement synthetic wavelength interferometry**. **(a)** A variant of the Michelson interferometer in Figure 2 that implements full-field synthetic wavelength interferometry. This setup produces wrong depth estimates because of the problems detailed in Figure 4. **(b)** A fiber scanning-based setup illuminates and images only one point on the scene, eliminating non-diagonal light paths such as the dashed ones in Figure 4 (a). Thus, getting rid of most global illumination, this setup produces more accurate depths than the one in (a). This comes at the price of needing to scan the imaged point on the scene, and doing so at a high-resolution megapixel-size grid is either impossible with current hardware or requires long acquisition time. **(c)** The interferometer in (a) can be made robust to global illumination by replacing the point emitter in the focal plane of the collimating lens in (a) with an area source. The area source is created by scanning a collimated laser beam and using a relay lens to focus the beam in the focal plane of the collimating lens.

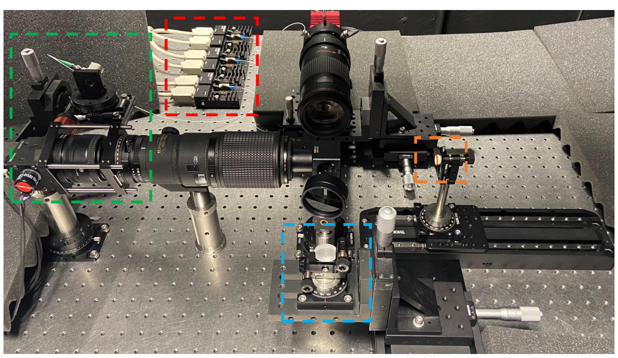


Fig. 6. **Picture of our hardware prototype**. Referring to Figure 5, **red**: combination of two single-mode lasers to create the illumination, **green**: swept-angle module, **blue**: reference arm consisting of a planar mirror on a translation stage, **orange**: scene.

In the next section, we introduce a new interferometer design, that combines the fast acquisition and high spatial resolution of full-field interferometers, with the robustness to indirect illumination of scanning interferometers.

4 SWEPT-ANGLE SYNTHETIC WAVELENGTH INTERFEROMETRY

In this section, we present our main contribution. We first show a solution to the global illumination problem illustrated in Figure 4 tweaking the interferometer as suggested by Kotwal et al. [2020]. Then, we show how to use this setup to take interferometric measurements and post-process them to estimate depth.

Spatially incoherent interferometry. Gkioulekas et al. [2015] showed that performing interferometry with spatially incoherent illumination *optically* rejects most global illumination from the interference measurement. They use in the Michelson interferometer in Figure 2 with a slight modification: Replacing the point emitter in the focal plane with an extended emitter such as an LED. This corresponds to replacing the $A(s) = \delta(s)$ in Figure 2 by $A(s) = \mathrm{rect}_a(s)$ for a source with spatial extent a. Kotwal et al. [2020] characterize the light paths that are permitted by such an extended emitter. We will use this result to state Proposition 1, which we prove in the supplement.

We model the interaction of light with the scene using its *complex light transmission function* $\mathcal{T}^c(x, x')$. The transmission function, a part of the Green's function for the scene, gives the amplitude and phase of the wave field generated at point x when a unit wave field originating at point x' interacts with the scene. Its discretized form can be viewed as the wave-optics analog of the familiar light transport formulation of O'Toole et al. [2012]. Of particular importance to us is the fact that $\mathcal{T}^c(x, x)$ measures the response of the scene at point x due to a source also located at x: This is equivalent to measuring the contributions of direct light paths such as the green ones in Figure 4(a) to scene appearance.

Proposition 1. *The complex correlation in Equation (3) when an emitter of size a with amplitude profile $A(s) = \mathrm{rect}_a(s)$ is placed in the focal plane of the collimating lens with focal length f equals:*

$$\boldsymbol{C}(x) = \frac{a}{f}\int_{x'} \mathcal{T}^c(x, x')\,\mathrm{sinc}\left(\frac{a\kappa\,(x - x')}{f}\right)\mathrm{d}x'. \tag{17}$$

As the size of the source a increases, the sinc function approaches a delta function. In that limit, Equation (17) says that the correlation equals $\mathcal{T}^c(x, x)$, which is the contribution of just direct light paths

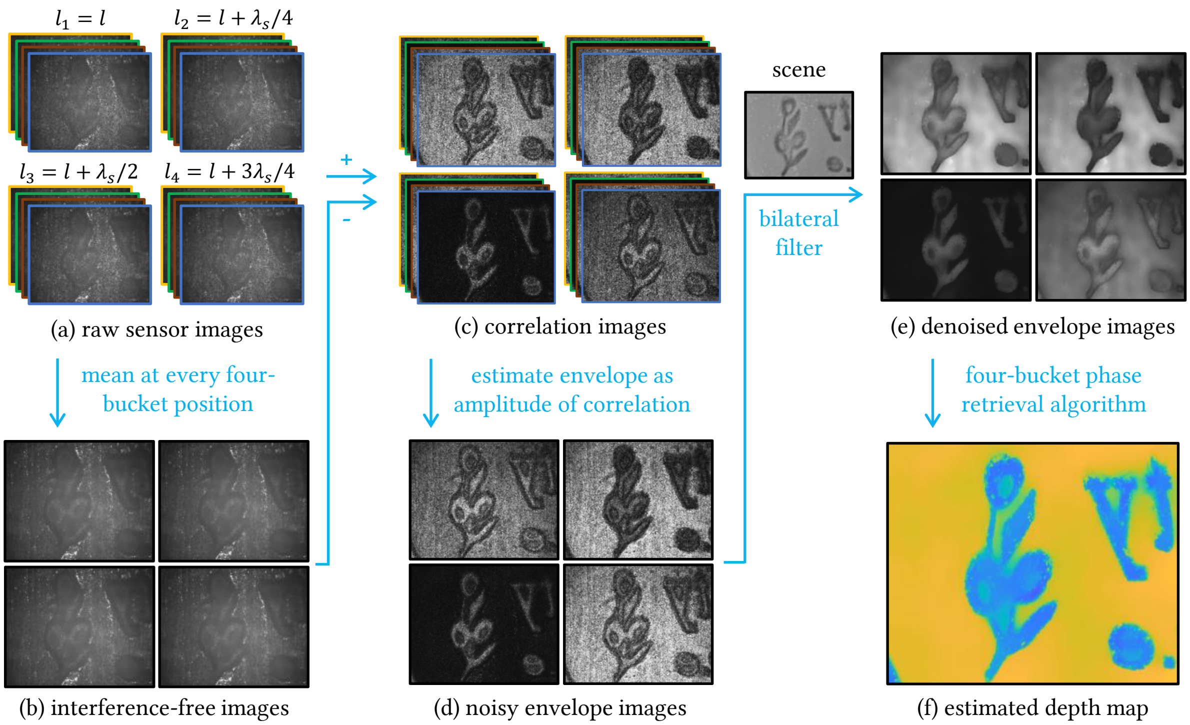


Fig. 7. **Acquisition pipeline for swept-angle synthetic wavelength interferometry**. **(a)** We take intensity measurements at sixteen positions of the reference arm: four *four-bucket positions* $l_n \equiv l + n\lambda_s/4$, and around every four-bucket position, four *sub-wavelength shifts* $\lambda_{mn} \equiv l_n + m\lambda_o/8$. **(b)** For every four-bucket position n, we estimate an *interference-free* image as the mean of its four subwavelength-shift images. **(c)** We subtract the interference-free images from the intensity images in (a) for every four-bucket position to estimate *interference-only* real parts of correlations Re $\{C(x, l_{mn})\}$. **(d)** At every four-bucket position, we estimate the envelope $\mathcal{E}(x, l_n)$ with the expression in Equation (19). **(e)** This estimate is noisy due to the presence of speckle, so we denoise it with a bilateral filter guided by an image of the scene taken under ambient light. Referring to Equation (15), the phase of the envelope sinusoid at a pixel x and reference arm position l_n is proportional to $d(x) - l_n$. The flat surface of the soap is placed perpendicular to the optical axis of the system, so it has the same depth throughout the scene. Therefore, the envelope sinusoid on the flat surface varies in unison across the l_n, making the entire flat surface completely dark or completely bright. **(f)** The four-bucket phase retrieval algorithm in Equation (22) yields the final depth.

to scene appearance. Thus, using a large enough source eliminates the contribution of the dashed paths in Figure 4(a) from the measured correlation C, and thus from the inferred envelope $\mathcal{E}$. As we will see further in Section 5, this results in a significant improvement in the quality of depth estimated for a variety of scenes.

Swept-angle interferometry. Combining the method of Gkioulekas et al. [2015] with synthetic wavelength interferometry, however, is challenging: Whereas Gkioulekas et al. [2015] used a *polychromatic* area source (such as a gas-discharge lamp), we need an area source emitting at two wavelengths. The extended light source closest to fulfilling this requirement is a sodium vapor lamp, emitting at the sodium D-lines around 589 nm [NIST 2013]. We found in experiments that emission at these lines in a sodium vapor lamp is not monochromatic enough for Equation (11) to hold. Also, the high temperatures and pressures used in these lamps often cause the lines to merge [van Bommel 2016], precluding their use in our setup.

We use Fourier-domain redistributive projector proposed by Kotwal et al. [2020] to create the required area source. As shown in the green dashed box in Figure 5(c), it takes in a collimated beam of light and uses a MEMS mirror to steer it. A relay lens then focuses this steered collimated beam in its focal plane, denoted by the dashed black line, thus as the MEMS steer light is scanned over an area, generating an area source. We then place this area source in the focal plane of the collimating lens that supplies light to the rest of the interferometer. Effectively, this arrangement performs time-division multiplexing of the input beam over the area of the required light source *within exposure*. This corresponds to sweeping through plane waves propagating at a set of directions at the output of the collimating lens. Due to this property, we term this kind of illumination as *swept-angle*, and call interferometry with such illumination *swept-angle interferometry*. Figure 6 shows a picture of the physical prototype of the swept-angle module in the green box.

Acquisition pipeline. The setup in Figure 5(c) gives us the ability to take swept-angle interferometric measurements at any desired position l of the reference arm. We will now show how to use the setup to estimate depth with swept-angle synthetic wavelength interferometry. As suggested by Equation (16), in order to estimate the scene depth $d(x)$, we need measurements of the envelope $\mathcal{E}(x, l)$ at four positions of the reference mirror: $l_n \equiv l + n\lambda_s/4$ for $n \in$

$\{0, 1, 2, 3\}$. The envelope, at every l_n, needs to be estimated from the real part of the correlation $\boldsymbol{C}(x, l)$. As Equation (14) says,

$$\mathrm{Re}\{\boldsymbol{C}(x, l)\} = 2\sin(2\kappa(d(x) - l))\,\boldsymbol{\mathcal{E}}(x, l) \tag{18}$$

Then, in order to estimate the envelope by itself, we take four measurements around every l_n, shifted by an eighth the optical wavelength λ_o: $l_{mn} \equiv l_n + m\lambda_o/8$ for $m \in \{0, 1, 2, 3\}$. Since the wavelength is orders of magnitude greater than the optical wavelength, shifts lower than the optical wavelength don't significantly change the phase of the envelope. We call these *sub-wavelength shifts*. Then, summing and squaring these, we get

$$\boldsymbol{\mathcal{E}}^2(x, l_n) = \frac{1}{8}\sum_{m=0}^{3}\left(\mathrm{Re}\{\boldsymbol{C}(x, l_{mn})\}\right)^2, \tag{19}$$

The square of the envelope estimated above can be written as

$$\boldsymbol{\mathcal{E}}^2(x, l_n) = \sin^2(\kappa\epsilon\,(d(x) - l)) \tag{20}$$

$$= \frac{1 - \cos(2\kappa\epsilon(d(x) - l))}{2} \tag{21}$$

Once the envelopes are known, then, the depth of the scene $d(x)$ can be estimated with the four-bucket algorithm as

$$d(x) = \frac{1}{2\kappa\epsilon}\arctan\left[\frac{\boldsymbol{\mathcal{E}}^2(x, l_3) - \boldsymbol{\mathcal{E}}^2(x, l_1)}{\boldsymbol{\mathcal{E}}^2(x, l_0) - \boldsymbol{\mathcal{E}}^2(x, l_2)}\right] + l + \frac{n\lambda_s}{2}. \tag{22}$$

However, we don't directly measure $\mathrm{Re}\{\boldsymbol{C}(x, l)\}$: Equation (2) tells us that it needs to be estimated from intensity measurements $\boldsymbol{I}(x, l_{mn})$ taken at positions l_{mn} of the reference arm. To estimate it, we first calculate at every four-bucket shift an *interference-free* image $\bar{\boldsymbol{I}}(x, l_n) \equiv \left(\sum_{m=0}^{3}\boldsymbol{I}(x, l_{mn})\right)/4$ as the mean of the frames at the sub-wavelength shifts. Then, $\mathrm{Re}\{\boldsymbol{C}(x, l_{mn})\} = \boldsymbol{I}(x, l_{mn}) - \bar{\boldsymbol{I}}(x, l_n)$. Figure 7 summarizes the acquisition pipeline with an example scene.

Dealing with speckle. Interference in diffuse scenes takes the form of *speckle*, a high-frequency pseudo-random pattern, as seen in Figure 7(d). Before speckle images can be used to estimate depth, they need to be denoised. The simplest denoising strategy is to convolve the speckle image with a low-pass filter (e.g. a Gaussian or a box filter). These filters, however, blur across fine features. To avoid this, we use a *bilateral* filter [Tomasi and Manduchi 1998] that performs denoising guided by an auxiliary image so as to not blur across fine features in the auxiliary image. We use as the auxiliary image an image of the scene taken under ambient light. We blur the envelopes estimated in Equation (15), and then use them to estimate depth using Equation (22).

We summarize the acquisition and post-processing steps in Algorithms 1 and 2 respectively.

Data: synthetic wavelength λ_s; optical wavelength λ_o; start position l
Result: intensity images $\boldsymbol{I}(x, l_{mn})$ at reference position l_{mn} (defined below)
$l_{mn} = l + n\lambda_s/4 + m\lambda_o/8$ for $n \in \{0, 1, 2, 3\}$ for $m \in \{0, 1, 2, 3\}$;
`/* Capture the intensity images in Figure 1 (a) */`
for *four-bucket positions* $n \in \{0, 1, 2, 3\}$ **do**
  for *sub-wavelength shifts* $m \in \{0, 1, 2, 3\}$ **do**
    move reference mirror to position l_{mn} ;
    capture image $\boldsymbol{I}(x, l_{mn})$;
  end
end
return $\boldsymbol{I}(x, l_{mn})$, $S(x)$

Algorithm 1: Acquiring intensity measurements with swept-angle synthetic wavelength interferometry

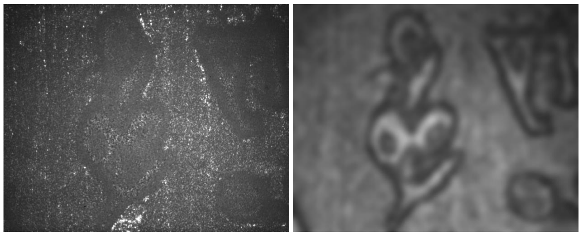
Data: synthetic wavelength λ_s; optical wavelength λ_o; start position l; bilateral filter hyperparameters: spatial kernel size σ_s and intensity kernel size σ_i; intensity measurements $\boldsymbol{I}(x, l_{mn})$ at reference position l_{mn} (defined below); scene ambient-light image $S(x)$
Result: depth map $d(x)$
`/* Initialization */`
$l_{mn} = l + n\lambda_s/4 + m\lambda_o/8$ for $n \in \{0, 1, 2, 3\}$ for $m \in \{0, 1, 2, 3\}$;
for *four-bucket positions* $n \in \{0, 1, 2, 3\}$ **do**
  `/* Figure 1 (b) */`
  estimate interference-free image
  $\bar{\boldsymbol{I}}(x, l_n) = \left(\sum_{m=0}^{3}\boldsymbol{I}(x, l_{mn})\right)/4$;
  `/* Figure 1 (c) */`
  estimate real parts of correlations
  $\mathrm{Re}\{\boldsymbol{C}(x, l_{mn})\} = \boldsymbol{I}(x, l_{mn}) - \bar{\boldsymbol{I}}(x, l_n)$;
  `/* Figure 1 (d) */`
  estimate noisy envelope
  $\tilde{\boldsymbol{\mathcal{E}}}^2(x, l_n) = \sum_{m=0}^{3}\left(\mathrm{Re}\{\boldsymbol{C}(x, l_{mn})\}\right)^2$;
  `/* Figure 1 (e) */`
  denoise envelope using the bilateral filter
  $\boldsymbol{\mathcal{E}}^2(x, l_n) = \mathrm{BilateralFilter}(\tilde{\boldsymbol{\mathcal{E}}}^2(x, l_n), S(x), \sigma_s, \sigma_i)$;
end
`/* Figure 1 (f) */`
estimate $d(x) = \frac{1}{2\kappa\epsilon}\arctan\left[\frac{\boldsymbol{\mathcal{E}}^2(x,l_3) - \boldsymbol{\mathcal{E}}^2(x,l_1)}{\boldsymbol{\mathcal{E}}^2(x,l_0) - \boldsymbol{\mathcal{E}}^2(x,l_2)}\right] + l$;
return $d(x)$

Algorithm 2: Processing intensity measurements to estimate depth in swept-angle synthetic wavelength interferometry

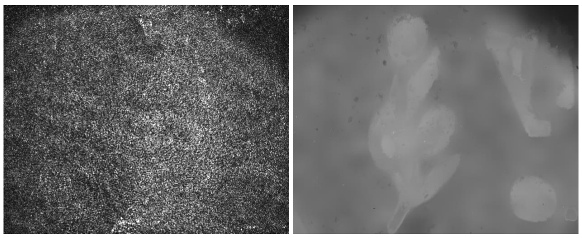
5 EXPERIMENTS

Depth recovery on challenging scenes. Here, we present examples of depth recovered using the prototype shown in Figure 6 on small scenes, about 1" in size. We divide our results into two figures based on the depth range of the scenes: Figure 9 with microscopic depth ranges (400 µm), and Figure 10 with macroscopic depth ranges (16 mm). These depth ranges are equal to the two synthetic wavelengths we used to avoid ambiguities in depth while reconstruction. The microscopic depth range was generated by combining two lasers as in Figure 5. We explain the creation of the macroscopic synthetic wavelengths in supplementary. In each category, we choose scenes challenging in various aspects:

(1) Material type: we choose scenes that go from metallic (coins in Figure 9) to diffuse (corner with two walls in Figure 10) to highly scattering (soap and chocolate in Figure 9 and toy cup in Figure 10);
(2) Spatial resolution: many of our scenes contain high-resolution spatial features packed together within tens of microns (music box and Harvard logo in Figure 9, twenty-dollar bill in Figure 1);



(a) with swept-angle source



(b) without swept-angle source

Fig. 8. **Importance of swept-angle scanning**. (a) Raw camera image (left) and extracted envelope (right) with swept-angle, and (b) without swept-angle. The swept-angle scanning rejects most global illumination from the interference component and keeps features with especially strong subsurface scattering, like the troughs of the engraving, intact.

(3) Depth resolution: many of our scenes require resolution of depth to the order of ten microns to accurately reconstruct fine depth details (Harvard logo in Figure 9, twenty-dollar bill in Figure 1).

For comparison, we also present depths captured with optical coherence tomography with the same setup by swapping the combination of two lasers out for a broadband light source. We captured OCT depths for microscopic scenes with a scan with step size 1 µm (thus requiring around 500 measurements), and those for macroscopic scenes with step size 10 µm (thus requiring around 1500 measurements). In comparison, our synthetic wavelength interferometry approach requires just sixteen measurements, making it tens to hundreds of times faster than OCT. Even with this speed-up and the complexity of the chosen scenes, our method with swept-angle probing and bilateral filtering recovers depth at spatial and depth resolutions comparable to, or in some cases better than OCT.

Figure 9 and Figure 10 also compare bilateral filtering against simple Gaussian filtering, showing the superior depth features achieved with bilateral filtering.

Importance of swept-angle scanning. Figure 9 and Figure 10 qualitatively show the importance of swept-angle scanning in synthetic wavelength interferometry. We see, especially in scenes with highly scattering light transport (like the chocolate and soap in Figure 9 and the toy cup in Figure 10), that the depth recovered without swept-angle is significantly worse than with, even unrecognizable at times. Below we report quantitative depth accuracy measurements, further confirming this observation.

To demonstrate the difference between measurements taken with and without swept-angle, in Figure 8 we visualize one of the raw data frames and one of the computed envelope images. It can be seen that the spatial incoherence introduced by the swept-angle scanning largely reduces interference speckle in the raw camera images. This reduction is due to swept-angle scanning rejecting global illumination coming from subsurface scattering in the soap. In addition, the envelope image computed with swept-angle scanning maintains the correct troughs of the engraving, where subsurface scattering is especially detrimental, while turning off the swept-angle scanning spatially blurs the signal.

Depth resolution. To verify our claim that our method recovers depth at micron-scale depth and spatial resolutions, we perform the following experiment. We place the chocolate scene from Figure 9 at different depths from the camera using a translation stage and capture depth with synthetic wavelength interferometry at each position. We chose this target due to its complex, sub-surface scattering nature. We perform this experiment with two parameters: with microscopic (λ_s = 400 µm) or macroscopic (λ_s = 16 mm) synthetic wavelengths, and with or without swept-angle interferometry. Figure 12 plots for all four experiments the recovered depth for a small, flat patch on the scene against the ground truth position of the scene translation stage with different Gaussian filter sizes. We see in Figure 12(a) and Figure 12(c) that with Gaussian filtering and swept-angle interferometry, with kernel sizes ≥ 11, we have strong positive correlation between the measured and ground truth depths. In contrast, in Figure 12(b) and Figure 12(d) there is no significant correlation between the measured and ground truth depths without swept-angle interferometry.

In Table 1 we additionally compute depth accuracy numbers. With Gaussian filtering and swept-angle, we see that, at a kernel size of 21 pixels (corresponding to a spatial extent of 140 µm), we estimate depth at microscopic scales to an accuracy of 1 µm and at macroscopic scales to an accuracy of 50 µm. The numbers for smaller and larger averaging windows quantify the trade-off between spatial resolution and depth accuracy. These numbers validate our claim that our method measures depth at micron-scale resolution. Note that in this case, as the target is planar we simply used a Gaussian filter for spatial averaging. In the presence of high frequency depth variations, this can be significantly improved using bilateral filtering, as demonstrated in Figure 9 and Figure 10.

Table 1. Quantitative evaluation of the resolution of our method. MAE is the mean absolute error between measured and ground truth depths, MedAE is the median absolute error and RMSE is the root mean square error. All errors are stated in µm.

type	kernel size	with swept-angle			without swept-angle		
		MAE	MedAE	RMSE	MAE	MedAE	RMSE
micro-	5 (35 µm)	6.5	4.8	8.2	15.2	13.2	18.9
scopic	11 (77 µm)	4.1	3.6	5.1	9.5	9.5	11.2
(period	15 (105 µm)	1.7	1.6	2.0	8.5	7.3	10.5
400 µm)	21 (147 µm)	1.3	**1.0**	1.6	8.6	6.7	11.1
macro-	5 (35 µm)	381.3	300.3	471.4	1130.4	1351.0	1267.2
scopic	11 (77 µm)	137.6	120.5	167.1	490.4	501.9	577.9
(period	15 (105 µm)	62.7	50.9	78.7	479.3	412.2	609.5
16 mm)	21 (147 µm)	60.9	**49.6**	81.7	484.9	334.4	605.7

Robustness to ambient light. In Figure 11, we demonstrate the robustness of our method to ambient light on the toy cup scene. We

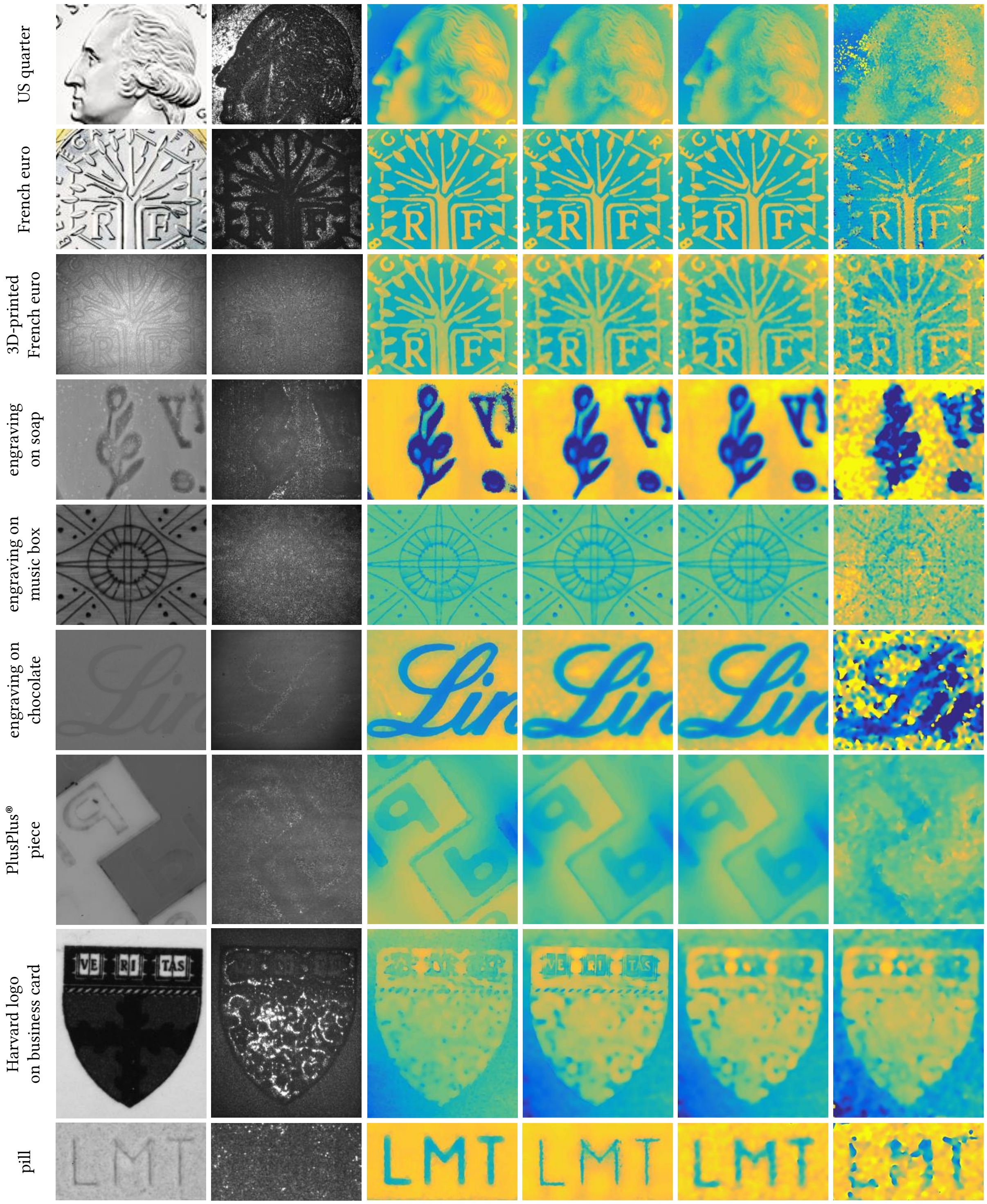


Fig. 9. **Depth reconstruction results for scenes with microscopic depth ranges**. From left to right: ambient light image, one of our raw images, depth measured using OCT, our depth with bilateral filtering and swept-angle, our depth with Gaussian filtering and swept-angle, our depth with Gaussian filtering and no swept-angle.

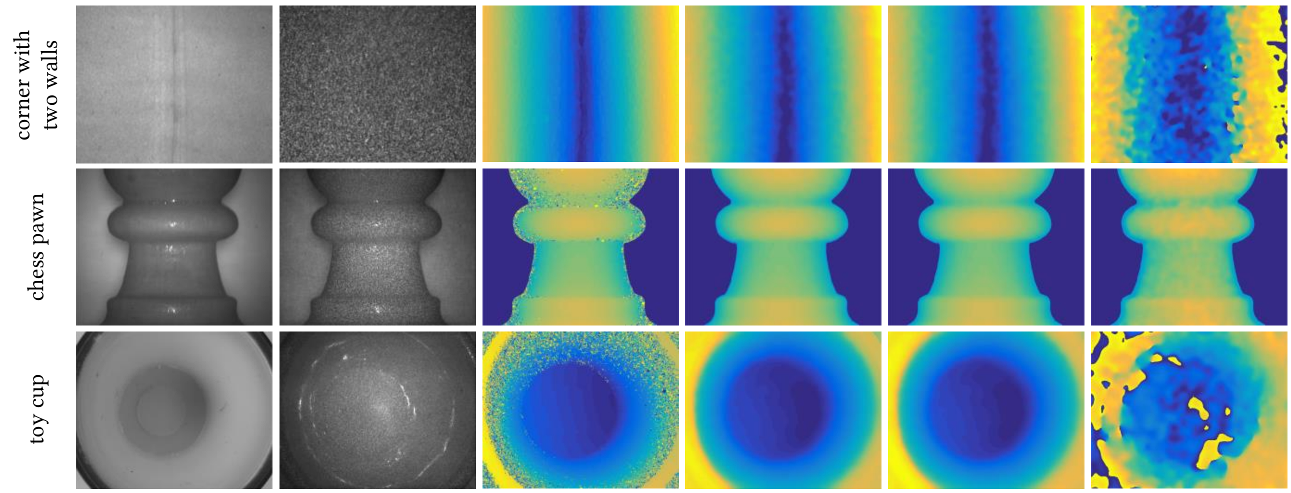
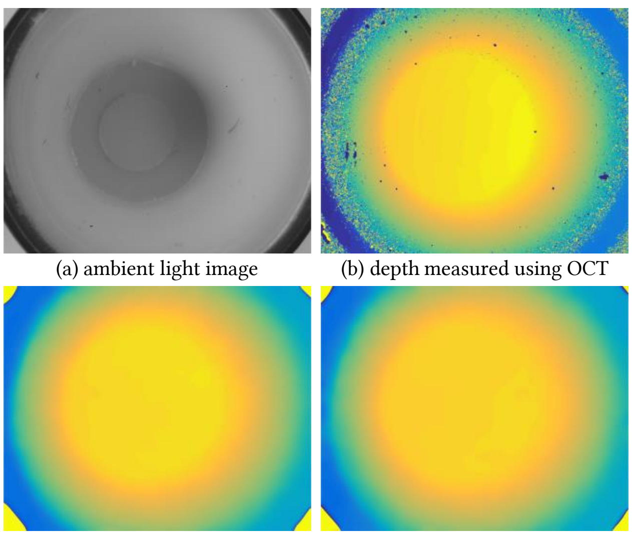


Fig. 10. **Depth reconstruction results for scenes with macroscopic depth ranges**. From left to right: ambient light image, one of our raw images, depth measured using OCT, our depth with bilateral filtering and swept-angle, our depth with Gaussian filtering and swept-angle, our depth with Gaussian filtering and no swept-angle.



(a) ambient light image (b) depth measured using OCT

(c) our depth without background (d) our depth with 10% SBR

Fig. 11. **Robustness of our method to external light**. In (d), we shine external light on the sample so that the signal-to-background ratio (SBR) our laser illumination to ambient noise is 0.1. This decreases the contrast of our interference speckle pattern, a sub-optimal situation for our method. Even in this situation, there is virtually no degradation in the quality of our recovered depth.

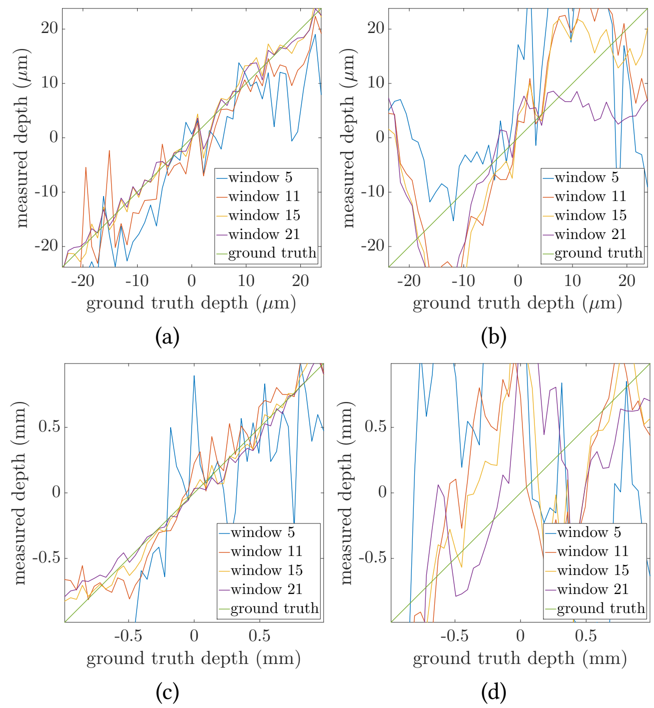
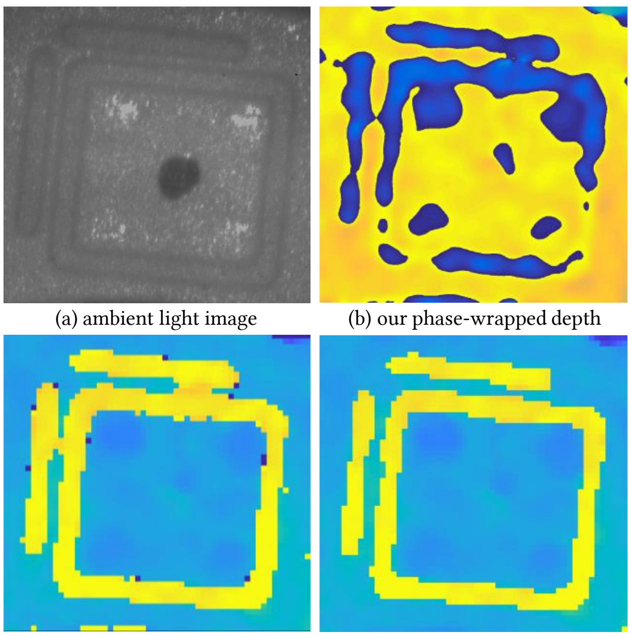


Fig. 12. **Testing the depth resolution of our method**. We place the chocolate scene from Figure 9 at different distances from the camera using a translation stage and capture measurements using our method at each position. We do this under four conditions: **(a)** microscopic synthetic wavelength with swept-angle, **(b)** microscopic synthetic wavelength without swept-angle, **(c)** macroscopic synthetic wavelength with swept-angle, and **(d)** macroscopic synthetic wavelength without swept-angle. In each case, we plot the depth measured by our method against the ground truth position of the scene provided by the translation stage. The window parameter in the plots is the size of the Gaussian blur kernel.

shine a spotlight on the scene such that the signal-to-background ratio (SBR) of the laser illumination to ambient light is 0.1. Ambient light adds to the intensity measurement at the camera, but not to interference, thus reducing interference contrast and potentially degrading the depth reconstruction. However, we see that at this SBR, the depth recovered from the toy cup scene (Figure 11 (d)) is virtually unchanged from the depth recovered without ambient light (Figure 11 (c)). In addition, to reject ambient light, we can use an ultra-narrow spectral filter centered at the illumination wavelength, an advantage OCT lacks because of its broadband spectrum.

6 DISCUSSION AND LIMITATIONS

Phase wrapping. As mentioned in Section 3, synthetic wavelength interferometry estimates depth up to a half-integer multiple of the synthetic wavelength λ_s: The depth that Equation 22 estimates lies in the range $[l - \lambda_s/4, l + \lambda_s/4]$. Any part of the scene that lies outside this range is 'wrapped' in the estimate to lie inside this range. We call this phenomenon 'phase wrapping'. We show an example of phase wrapping in Figure 13, where we try to estimate the depth of a pattern on a poker chip with a synthetic wavelength $\lambda_s = 170\,\mu$m. To unwrap the phase it is common to use measurements at multiple



(a) ambient light image (b) our phase-wrapped depth (c) unwrapped depth without regularization (d) unwrapped depth with regularization

Fig. 13. **Phase unwrapping**. When the depth range of the scene is greater than the synthetic wavelength, phase-wrapping artifacts, such as those in (b), result. To recover depth free of wrapping, we can repeat our measurement procedure with multiple different synthetic wavelengths and solve a phase unwrapping problem resulting in depth (c). With some spatial regularization of the estimated depth we can clean the estimate of the artifacts somewhat, resulting in (d). This method is applied on sub-sampled images for reasonable running times.

synthetic wavelengths [Cheng and Wyant 1984, 1985; Droeschel et al. 2010]. To try this approach we took interferometric measurements for the poker chip scene in Figure 13 with synthetic wavelengths 170 µm, 300 µm and 400 µm. With these measurements, we solved a spatially-regularized inverse problem with loopy belief propagation to estimate the depth. For reasonable running times, we downsampled our measurements to 5% of their original size. With a running time of 100 s, we estimate the depth map shown in Figure 13(d). We acknowledge that the optimization is highly non-convex, and this proof-of-concept low-resolution result required very careful parameter tuning. Therefore, we leave further development of phase unwrapping for our method to future work.

Comparison with full-field optical coherence tomography. As we mentioned earlier, optical coherence tomography is another interferometric modality that can be used for micron-scale depth acquisition. It can be implemented with a similar Michelson interferometer with swept-angle scanning [Gkioulekas et al. 2015], but requires and exhaustive scan of the reference mirror throughout the depth range, resulting in very long acquisition time. As opposed to this, synthetic wavelength interferometry requires us to translate the reference mirror to only sixteen positions, resulting in a huge speed-up and increased tolerance to vibrations over OCT. In addition, synthetic wavelength interferometry has the advantage that it uses ultra-narrow-bandwidth illumination, allowing us to reject ambient light much better than OCT by using a narrow spectral filter.

However, this speed-up comes at the price of *coupled depth range and axial resolution.* The depth range of OCT is restricted only by reference mirror translation range, while the axial resolution is limited only by the width of $C_{\mathrm{OCT}}(x, l)$ around $d(x)$ and reference mirror translation resolution. These are independent properties of the implementation, and therefore can be changed without affecting each other. As opposed to this, the depth range of swept-angle synthetic wavelength interferometry is coupled with axial resolution, and there is no clear way of trading one off for the other. In addition, the coherent illumination used in swept-angle synthetic wavelength interferometry makes the captured images very susceptible to speckle noise. Also, due to the precise relation between the synthetic wavelength and individual laser wavelengths, the synthetic wavelength is very sensitive to laser operating conditions.

Comparison with scanning-based interferometric techniques. Both our swept-angle technique and impulse projection reject global illumination using scanning. Whereas swept-angle scans are performed in the focal plane of the collimating lens, impulse projection systems scan across the scene. Both kinds of beam steering can be implemented by using fast MEMS mirrors: However, in the case of an impulse projection system, the sensor needs to be synchronized with the micromirror to map the position of the mirror to the sensor measurement. Then, the exposure time at each spatial position needs to be high enough to collect enough photons, significantly slowing down acquisition. In contrast, swept-angle scanning happens within exposure, causing no such time penalty.

In addition, the spatial resolution of an impulse projection system is determined by the spot size of the beam on the scene. A broad illuminated spot causes light from neighboring points to 'leak' into the imaged point, thus imaging a spatially blurred version of the true depth. As opposed to this, swept-angle scanning allows us to capture images with resolutions as high as the sensor's pixel pitch.

Direct versus diagonal light paths. Our approach effectively acquires the diagonal of the light transport matrix. This includes the direct path but also a small amount of non direct paths starting and ending at the same scene point (see yellow path in Figure 4). In most scenes non-direct diagonal light paths are negligible relative to the direct ones. However, there do exist special circumstances when non-direct diagonal light paths contribute significantly to scene appearance. These scenes are impossible to reconstruct accurately with synthetic wavelength interferometry without strategies like the ones proposed by Kadambi et al. [2013] and Fuchs [2010].

Real-time operation. The hardware prototype in Figure 6 can currently acquire measurements at a frame rate of 1 Hz. The bottleneck to real-time operation is performing the sixteen shifts by physically translating the reference mirror. Acquiring a translation stage with faster travel speeds will allow us to increase this frame rate.

One direction towards making the shifts faster is to perform the sub-wavelength shifts rapidly by replacing the reference mirror with a free-space phase spatial light modulator (as done by Kotwal et al. [2020]). However, the four-bucket shifts of the synthetic wavelength, being out of the range of achievable phase shifts for the modulator, would still need to be implemented with physical movement. A potential replacement for the movement could be using space-division multiplexing, by introducing a custom phase mask implementing a repeating 2×2 pattern of $\{0, \lambda_s/4, \lambda_s/2, 3\lambda_s/4\}$ optical path length shifts in the reference arm light path. Alternatively, space-division multiplexing can be implemented with polarization analyzers, with different phase shifts applied to each polarization [Maeda et al. 2018]. Space-division multiplexing comes at the cost of spatial resolution, which can reduce the quality of depth estimates. Another potential solution would be to use an arrangement of four mirrors spaced at $\lambda_s/4$ to switch between, eliminating the translation.

7 CONCLUSION

We presented a method for fast, micron-scale spatial- and axial-resolution depth sensing with swept-angle synthetic wavelength interferometry. This method uses light consisting of two narrowly-separated optical wavelengths. We showed that the envelope of per-pixel interferometric measurements in such a setup is a sinusoid that depends inversely upon the separation between the wavelengths, and whose phase encodes the depth of the pixel. We added a swept-angle scanning mechanism that provides pixel-level robustness to global illumination. Finally, we proposed filtering interferometric measurements with an ambient light image of the scene to preserve high-resolution features in the estimated depth.

We tested this theory by capturing high-quality depth for a variety of scenes with fine spatial features, facing challenges from complicated light transport effects such as interreflections, subsurface scattering and specularities. We validated our claim of micron axial resolution by measuring numerical accuracy of estimated depth, and showed the robustness of our method to ambient illumination.

We hope that the quality of our results will inspire applications of swept-angle synthetic wavelength interferometry in applications such as high-resolution computer vision, biomedical imaging and industrial fabrication. In addition, we hope that the utility we have demonstrated will lead to future research on phase unwrapping and robust reconstruction with interferometry data.

REFERENCES

Nils Abramson. 1983. Light-in-flight recording: high-speed holographic motion pictures of ultrafast phenomena. *Applied Optics* (1983).

Supreeth Achar, Joseph R Bartels, William L Whittaker, Kiriakos N Kutulakos, and Srinivasa G Narasimhan. 2017. Epipolar time-of-flight imaging. *ACM TOG* (2017).

Brian Aull. 2005. 3D Imaging with Geiger-mode Avalanche Photodiodes. *Opt. Photon. News* (2005).

Stephen T. Barnard and William B. Thompson. 1980. Disparity Analysis of Images. *IEEE TPAMI* (1980).

Ayush Bhandari, Achuta Kadambi, Refael Whyte, Christopher Barsi, Micha Feigin, Adrian Dorrington, and Ramesh Raskar. 2014. Resolving multipath interference in time-of-flight imaging via modulation frequency diversity and sparse regularization. *Optics Letters* (2014).

John H. Bruning, Donald R. Herriott, Joseph E. Gallagher, Daniel P. Rosenfeld, Andrew D. White, and Donald J. Brangaccio. 1974. Digital Wavefront Measuring Interferometer for Testing Optical Surfaces and Lenses. *Applied Optics* (1974).

Tongbo Chen, Hans-Peter Seidel, and Hendrik P. A. Lensch. 2008. Modulated phase-shifting for 3D scanning. In *IEEE/CVF CVPR*.

Yeou-Yen Cheng and James C. Wyant. 1984. Two-wavelength phase shifting interferometry. *Applied Optics* (1984).

Yeou-Yen Cheng and James C. Wyant. 1985. Multiple-wavelength phase-shifting interferometry. *Applied Optics* (1985).

Katherine Creath. 1985. Phase-shifting speckle interferometry. *Applied Optics* (1985).

Peter de Groot. 2011. *Phase Shifting Interferometry*.

Peter de Groot and John McGarvey. 1992. Chirped synthetic-wavelength interferometry. *Optics Letters* (1992).

David Droeschel, Dirk Holz, and Sven Behnke. 2010. Multi-frequency Phase Unwrapping for Time-of-Flight cameras. In *IEEE/RSJ IROS*.

Adolf F. Fercher, H Z. Hu, and U. Vry. 1985. Rough surface interferometry with a two-wavelength heterodyne speckle interferometer. *Applied Optics* (1985).

Richard Ferriere, Johann Cussey, and John M. Dudley. 2008. Time-of-flight range detection using low-frequency intensity modulation of a cw laser diode: application to fiber length measurement. *Optical Engineering* (2008).

Angel Flores, Craig Robin, Ann Lanari, and Iyad Dajani. 2014. Pseudo-random binary sequence phase modulation for narrow linewidth, kilowatt, monolithic fiber amplifiers. *Optics Express* (2014).

Daniel Freedman, Eyal Krupka, Yoni Smolin, Ido Leichter, and Mirko Schmidt. 2014. SRA: Fast Removal of General Multipath for ToF Sensors. arXiv:1403.5919 [cs.CV]

Stefan Fuchs. 2010. Multipath Interference Compensation in Time-of-Flight Camera Images. In *ICPR*.

Genevieve Gariepy, Nikola Krstajić, Robert Henderson, Chunyong Li, Robert R Thomson, Gerald S Buller, Barmak Heshmat, Ramesh Raskar, Jonathan Leach, and Daniele Faccio. 2015. Single-photon sensitive light-in-fight imaging. *Nature Comm.* (2015).

Ioannis Gkioulekas, Anat Levin, Frédo Durand, and Todd Zickler. 2015. Micron-scale light transport decomposition using interferometry. *ACM TOG* (2015).

P. Grossmann. 1987. Depth from focus. *PRL* (1987).

Anant Gupta, Atul Ingle, and Mohit Gupta. 2019a. Asynchronous Single-Photon 3D Imaging. In *ICCV*.

Anant Gupta, Atul Ingle, Andreas Velten, and Mohit Gupta. 2019b. Photon-Flooded Single-Photon 3D Cameras. In *IEEE/CVF CVPR*.

Mohit Gupta, Amit Agrawal, Ashok Veeraraghavan, and Srinivasa G Narasimhan. 2011. Structured light 3D scanning in the presence of global illumination. In *IEEE/CVF CVPR*.

Mohit Gupta, Shree K Nayar, Matthias B Hullin, and Jaime Martin. 2015. Phasor imaging: A generalization of correlation-based time-of-flight imaging. *ACM TOG* (2015).

Mohit Gupta, Andreas Velten, Shree K. Nayar, and Eric Breitbach. 2018. What Are Optimal Coding Functions for Time-of-Flight Imaging? *ACM TOG* (2018).

Felipe Gutierrez-Barragan, Syed Azer Reza, Andreas Velten, and Mohit Gupta. 2019. Practical Coding Function Design for Time-Of-Flight Imaging. In *2019 IEEE/CVF CVPR*.

Yudeog Han, Joon-Young Lee, and In So Kweon. 2013. High Quality Shape from a Single RGB-D Image under Uncalibrated Natural Illumination. In *IEEE ICCV*.

Parameswaran Hariharan. 2003. *Optical interferometry*. Elsevier.

Richard Hartley and Andrew Zisserman. 2004. *Multiple View Geometry in Computer Vision*.

Caner Hazirbas, Sebastian Georg Soyer, Maximilian Christian Staab, Laura Leal-Taixé, and Daniel Cremers. 2018. Deep Depth From Focus. arXiv:1704.01085 [cs.CV]

Felix Heide, Steven Diamond, David B. Lindell, and Gordon Wetzstein. 2018. Sub-picosecond photon-efficient 3D imaging using single-photon sensors. *Scientific Reports* (2018).

Felix Heide, Matthias B Hullin, James Gregson, and Wolfgang Heidrich. 2013. Low-budget transient imaging using photonic mixer devices. *ACM TOG* (2013).

Berthold Klaus Paul Horn. 1970. *Shape from Shading: A Method for Obtaining the Shape of a Smooth Opaque Object from One View.* Technical Report.

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, and James G Fujimoto. 1991. Optical coherence tomography. *Science* (1991).

David Jimeneza, Daniel Pizarrob, Manuel Mazoa, and Sira Palazuelos. 2014. Modeling and correction of multipath interference in time of flight cameras. *Image and Vision Computing* (2014).

Jon L. Johnson, Timothy D. Dorney, and Daniel M. Mittleman. 2001. Enhanced depth resolution in terahertz imaging using phase-shift interferometry. *Applied Physics Letters* (2001).

Achuta Kadambi, Refael Whyte, Ayush Bhandari, Lee Streeter, Christopher Barsi, Adrian Dorrington, and Ramesh Raskar. 2013. Coded Time of Flight Cameras: Sparse Deconvolution to Address Multipath Interference and Recover Time Profiles. *ACM TOG* (2013).

Ahmed Kirmani, Tyler Hutchison, James Davis, and Ramesh Raskar. 2009. Looking around the corner using transient imaging. In *IEEE/CVF CVPR*.

Ahmed Kirmani, Dheera Venkatraman, Dongeek Shin, Andrea Colaço, Franco N. C. Wong, Jeffrey H. Shapiro, and Vivek K Goyal. 2014. First-Photon Imaging. *Science* (2014).

Christ Leonidas Koliopoulos. 1981. *Interferometric Optical Phase Measurement Techniques.* Ph.D. Dissertation. THE UNIVERSITY OF ARIZONA.

Alankar Kotwal, Anat Levin, and Ioannis Gkioulekas. 2020. Interferometric transmission probing with coded mutual intensity. *ACM TOG* (2020).

Robert Lange and Peter Seitz. 2001. Solid-state time-of-flight range camera. *IEEE JQE* (2001).

Robert Lange, Peter Seitz, Alice Biber, and Stefan Lauxtermann. 2000. Demodulation pixels in CCD and CMOS technologies. *SPIE* (2000).

Fengqiang Li, Florian Willomitzer, Prasanna Rangarajan, Mohit Gupta, Andreas Velten, and Oliver Cossairt. 2018. Sh-tof: Micro resolution time-of-flight imaging with superheterodyne interferometry. *IEEE ICCP* (2018).

Fengqiang Li, Joshua Yablon, Andreas Velten, Mohit Gupta, and Oliver Cossairt. 2017. High-depth-resolution range imaging with multiple-wavelength superheterodyne interferometry using 1550-nm lasers. *Applied Optics* (2017).

David B Lindell, Matthew O'Toole, and Gordon Wetzstein. 2018. Single-photon 3D imaging with deep sensor fusion. *ACM TOG* (2018).

Xiaomeng Liu, Kristofer Henderson, Joshua Rego, Suren Jayasuriya, and Sanjeev Koppal. 2021. Dense Lissajous sampling and interpolation for dynamic light-transport. *Optics Express* (2021).

Tomohiro Maeda, Achuta Kadambi, Yoav Y Schechner, and Ramesh Raskar. 2018. Dynamic heterodyne interferometry. *IEEE ICCP* (2018).

Julio Marco, Quercus Hernandez, Adolfo Muñoz, Yue Dong, Adrian Jarabo, Min H. Kim, Xin Tong, and Diego Gutierrez. 2017. DeepToF: Off-the-Shelf Real-Time Correction of Multipath Interference in Time-of-Flight Imaging. *ACM TOG* (2017).

Karl Meiners-Hagen, René Schödel, Florian Pollinger, and Ahmed Abou-Zeid. 2009. Multi-Wavelength Interferometry for Length Measurements Using Diode Lasers.

Inc. Mirrorcle Technologies. 2022. Mirrorcle MEMS Technical Overview. https://www.mirrorcletech.com/pdf/Mirrorcle_MEMS_Mirrors_-_Technical_Overview.pdf.

Nikhil Naik, Achuta Kadambi, Christoph Rhemann, Shahram Izadi, Ramesh Raskar, and Sing Bing Kang. 2015. A light transport model for mitigating multipath interference in Time-of-flight sensors. In *IEEE/CVF CVPR*.

Lazaros Nalpantidis, Georgios Sirakoulis, and Antonios Gasteratos. 2008. Review of Stereo Vision Algorithms: From Software to Hardware. *Int. J. Optomechatronics* (2008).

Shree Nayar, Gurunandan Krishnan, Michael Grossberg, and Ramesh Raskar. 2006. Fast separation of direct and global components of a scene using high frequency illumination. *ACM TOG* (2006).

Cristiano Niclass, Alexis Rochas, Pierre. Besse, and Edoardo Charbon. 2005. Design and characterization of a CMOS 3-D image sensor based on single photon avalanche diodes. *IEEE JSSC* (2005).

NIST. 2013. *Handbook of Basic Atomic Spectroscopic Data*. Technical Report.

Matthew O'Toole, Felix Heide, David B Lindell, Kai Zang, Steven Diamond, and Gordon Wetzstein. 2017. Reconstructing Transient Images from Single-Photon Sensors. *IEEE/CVF CVPR* (2017).

Matthew O'Toole, Felix Heide, Lei Xiao, Matthias B Hullin, Wolfgang Heidrich, and Kiriakos N Kutulakos. 2014a. Temporal Frequency Probing for 5D Transient Analysis of Global Light Transport. *ACM TOG* (2014).

Matthew O'Toole, John Mather, and Kiriakos N Kutulakos. 2014b. 3D Shape and Indirect Appearance by Structured Light Transport. *IEEE/CVF CVPR* (2014).

Matthew O'Toole, John Mather, and Kiriakos N. Kutulakos. 2016. 3D Shape and Indirect Appearance by Structured Light Transport. *IEEE TPAMI* (2016).

Matthew O'Toole, Ramesh Raskar, and Kiriakos N Kutulakos. 2012. Primal-dual Coding to Probe Light Transport. *ACM TOG* (2012).

Andrew Payne, A Dorrington, and Michael Cree. 2011. Illumination Waveform Optimization for Time-of-Flight Range Imaging Cameras. *SPIE* (2011).

Dario Piatti, Fabio Remondino, and David Stoppa. 2013. *State-of-the-Art of TOF Range-Imaging Sensors.*

Dikpal Reddy, Ravi Ramamoorthi, and Brian Curless. 2012. Frequency-space Decomposition and Acquisition of Light Transport Under Spatially Varying Illumination. *IEEE ECCV* (2012).

Alexis Rochas, Michael Gosch, Alexandre Serov, Pierre Besse, Rade S. Popovic, T. Lasser, and Rudolph Rigler. 2003. First fully integrated 2-D array of single-photon detectors in standard CMOS technology. *IEEE PTL* (2003).

Daniel Scharstein and Richard Szeliski. 2003. High-accuracy stereo depth maps using structured light. In *IEEE/CVF CVPR*.

Rudolf Schwarte, Zhanping Xu, Horst-Guenther Heinol, Joachim Olk, Ruediger Klein, Bernd Buxbaum, Helmut Fischer, and Juergen Schulte. 1997. New electro-optical mixing and correlating sensor: facilities and applications of the photonic mixer device (PMD). In *Sensors, Sensor Systems, and Sensor Data Processing*.

Murali Subbarao and Gopal Surya. 1994. Depth from defocus: A spatial domain approach. *IJCV* (1994).

Orazio Svelto. 2010. *Principles of Lasers*.

Thorlabs, Inc. 2022. F-Theta Scan Lenses. https://www.thorlabs.com/newgrouppage9.cfm?objectgroup_id=6430.

Carlo Tomasi and Roberto Manduchi. 1998. Bilateral filtering for gray and color images. In *IEEE ICCV*.

Wout van Bommel. 2016. *High- and Low-Pressure Sodium Lamp*.

Andreas Velten, Thomas Willwacher, Otkrist Gupta, Ashok Veeraraghavan, Moungi G. Bawendi, and Ramesh Raskar. 2012. Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging. *Nature Comm.* (2012).

Federica Villa, Rudi Lussana, Danilo Bronzi, Simone Tisa, Alberto Tosi, Franco Zappa, Alberto Dalla Mora, Davide Contini, Daniel Durini, Sasha Weyers, and Werner Brockherde. 2014. CMOS Imager With 1024 SPADs and TDCs for Single-Photon Timing and 3-D Time-of-Flight. *IEEE JSTQE* (2014).

James C. Wyant, Chris L. Koliopoulos, Bharat Bhushan, and Orrin E. George. 1984. An Optical Profilometer for Surface Characterization of Magnetic Media. *ASLE Transactions* (1984).

A PROOF OF PROPOSITION 1

Proposition 1. *The complex correlation in Equation 3 when an emitter of size a with a flat amplitude profile $A(s) = \mathrm{rect}_a(s)$ is placed in the focal plane of the collimating lens with focal length f equals:*

$$\boldsymbol{C}(x) = \frac{a}{f} \int_{x'} \mathcal{T}^c\left(x, x'\right) \operatorname{sinc}\left(\frac{a\kappa\left(x - x'\right)}{f}\right) \mathrm{d}x'. \tag{23}$$

Proof. We re-state the result from Kotwal et al. [2020] here. We make the paraxial assumption, so that at all non-zero points s on the source, $|s| < a \ll f$. Then, reparametrizing the source coordinates in Kotwal et al. [2020, Equation (31)] as $\theta \equiv s/f$, the correlation $\boldsymbol{C}(x, l)$ equals the convolution of the transmission function with the *probing kernel* $\boldsymbol{\mathcal{P}}^c$

$$\boldsymbol{C}(x, l) = \int_{x'} \mathcal{T}^c\left(x, x'\right) \boldsymbol{\mathcal{P}}^c\left(x - x'\right) \mathrm{d}x', \tag{24}$$

where $\boldsymbol{\mathcal{P}}^c$ is expressed in terms of $A(s)$ as

$$\boldsymbol{\mathcal{P}}^c(x - x') = \frac{1}{f} \int_s A(s) \exp(-i\kappa s(x - x')/f)\, \mathrm{d}s. \tag{25}$$

We will first calculate the probing kernel.

$$\begin{aligned}\boldsymbol{\mathcal{P}}^c(x - x') &= \frac{1}{f} \int_s A\left(s\right) \exp\left(-i\kappa s\left(x - x'\right)/f\right)\, \mathrm{d}s \\ &= \frac{1}{f} \int_s \mathrm{rect}_a\left(s\right) \exp\left(-i\kappa s\left(x - x'\right)/f\right)\, \mathrm{d}s \\ &= \frac{a}{f} \operatorname{sinc}\left(\frac{a\kappa\left(x - x'\right)}{f}\right)\end{aligned} \tag{26}$$

Then, following Kotwal et al. [2020, Equation (31)], the correlation is given by

$$\begin{aligned}\boldsymbol{C}(x) &= \int_{x'} \mathcal{T}^c\left(x, x'\right) \boldsymbol{\mathcal{P}}^c\left(x - x'\right) \mathrm{d}x' \\ &= \frac{a}{f} \int_{x'} \mathcal{T}^c\left(x, x'\right) \operatorname{sinc}\left(\frac{a\kappa\left(x - x'\right)}{f}\right) \mathrm{d}x'\end{aligned} \tag{27}$$

This concludes the proof. □

B DIFFICULTIES IN FAST IMAGE-DOMAIN SCANNING

In this section, we discuss our claim at the end of Section 3 in the main paper. We mentioned that the robustness of scanning interferometers to global illumination comes at the cost of having to use beam steering to scan the entire scene. We will elaborate on the three requirements for a scanning inteferometer imaging at micrometer-scale axial *and* spatial resolutions: (i) a laser beam a few micrometers wide; (ii) a MEMS mirror capable of scanning at high-enough angular resolution to translate the laser beam a few microns on the scene surface; and (iii) acquisition time long enough to scan a megapixel-size grid on the scene. Here is why each of them, respectively, is difficult to meet:

(i) The diameter of a Gaussian laser beam is inversely proportional to its divergence [Svelto 2010, Chapter 4]. The smaller the beam diameter, the larger the divergence: Therefore, maintaining a collimated micron-diameter laser beam is difficult. At 780 nm, a laser beam with a diameter of 1 micron grows in diameter by 10% every 2 m.

Alternatively to thin, collimated laser beams, we can also *focus* the output of the fiber in Figure 5(b) onto the scene surface. Contrary to micron-scale beam waists, it is possible to focus pump lasers to spot sizes of tens of microns [Svelto 2010, Chapter 9]: for example, Thorlabs manufactures $f - \theta$ scanning lenses that yield diffraction-limited spot sizes of a minimum of 15 µm at 1046 nm [Thorlabs, Inc. 2022]. However, focusing the laser beam onto the scene sharply decreases the depth of field of the imaging system from being limited by the divergence of the collimated beam to being limited by the quadratic phase profile of the focused spot. In order to use this focused setup, then, we need another scan over the position of the focused spot, which only adds to acquisition time.

(ii) Top-of-the-line scanning micromirrors typically have angular scanning resolutions of 10 µrad [Mirrorcle Technologies 2022]. The maximum distance to the scene such that it can be scanned at micron spatial resolution is then 10 cm.

(iii) The scanning micromirror needs to be run in 'point-to-point scanning mode' [Mirrorcle Technologies 2022] where the micromirror stops at every desired position. The best settling times for step mirror deflections are around 100 µs [Mirrorcle Technologies 2022]. Using these numbers, for a megapixel image, just micromirror rotations take up 100 s.

A full-field interferometer does not need any scanning over the scene. Instead, it accomplishes imaging diagonal paths by scanning an area source in the focal plane of the collimating lens, an operation that can be done in the resonant mode of a MEMS mirror within exposure.

C ACQUISITION SETUP

We discuss here the engineering details of the setup implementing synthetic wavelength interferometry. The schematic and a picture of the setup are shown in Figure 5 (c) of the main paper. We largely use the same components as in the setup of Kotwal et al. [2020], and replicate the implementation details below for completeness.

Light source. We use near-infrared single frequency tunable laser diodes from Thorlabs (DBR780PN, 780 nm, 45 mW, 1 MHz linewidth). These laser diodes are tunable in wavelength by adjusting either operating current or temperature of the diode. To create small wavelength separations (of the order of 0.01 nm), we modulate the operating current of one laser diode with a square waveform, thus create two time-multiplexed wavelengths. To create larger separations (of the order of 1 nm), we use two different laser diodes selected at the appropriate central wavelengths. This is possible because the central wavelengths of separately manufactured laser diodes vary in a ±2 nm region around 780 nm. We found that for accurate depth recovery, it is important for the light sources used to be monochromatic (single longitudinal mode), stable in wavelength and power, and accurately tunable. We experimented with multiple alternatives and encountered problems with either stability, tunability or monochromaticity. We found the DBR lasers from Thorlabs optimal in all these aspects.

Table 2. List of major components used in the optical setup of Figure 5 (c) of the main paper.

description	quantity	model name	company
single-frequency lasers, 780 nm CWL, 45 mW power	2	DBR780PN	Thorlabs
benchtop laser diode current controller, ± 500 mA HV	2	LDC205C	Thorlabs
benchtop temperature controller, ± 2 A / 12 WW	2	TED200C	Thorlabs
1×2 polarization-maintaining fiber coupler, 780 ± 15 nm	1	PN780R5A1	Thorlabs
reflective FC/APC fiber collimator	1	RC04APC-P01	Thorlabs
2× beam expander	1	GBE02-B	Thorlabs
2-axis galvanometer mirror set	1	GVS202	Thorlabs
function generator	2	SDG1025	Siglent
35 mm compound lens	1	AF Micro Nikkor 35mm 1:4 D IF-ED	Nikon
200 mm compound lens	1	AF Micro Nikkor 200mm 1:4 D IF-ED	Nikon
25 mm × 36 mm plate beamsplitter	3	BSW10R	Thorlabs
1 inch round protected Aluminum mirror	3	ME1-G01	Thorlabs
2 inch absorptive neutral density filter kit	1	NEK01S	Thorlabs
ultra-precision linear motor stage, 16 cm travel	1	XMS160	Newport Corporation
ethernet driver for linear stage	1	XPS-Q2	Newport Corporation
780.5 ± 1 nm OD6 ultra-narrow spectral filter	1	-	Alluxa
180 mm compound lens	1	EF 180mm f/3.5L Macro USM	Canon
8 MP CCD color camera with Birger EF mount	1	PRO-GT3400-09	Allied Vision Technologies

Estimating the synthetic wavelength. The synthetic wavelength resulting from this illumination is very sensitive to the separation between the two wavelengths, especially at microscopic scales. Therefore, after selecting a pair of lasers or current levels for an approximate synthetic wavelength, it is necessary to estimate the actual synthetic wavelength accurately. To do this, we measure the magnitudes of the complex interference patterns at a dense collection of reference arm positions. We then fit a sinusoid to these measured magnitudes and use the fit wavelength as the synthetic wavelength. In practice, we have a series of measurements at each pixel, so we use the median of the wavelength estimates at all pixels.

Mechanism for imaging direct-only photon paths. We use two fast-rotating mirrors to scan the incoming collimated laser beam in a square $1^\circ \times 1^\circ$ angular pattern at kHz frequencies, as shown by Kotwal et al. [2020] and Liu et al. [2021]. An intermediate lens (a 35 mm Nikon prime lens) then maps angle into spatial position behind the illumination lens, creating the 'source'. The mirrors are operated by a function generator generating sinusoids at kHz frequency with a slight offset to create a dense Lissajous curve that spans a square. As mentioned in the above sections and by Kotwal et al. [2020], measuring interference with such a light source results in imaging only direct photon paths. Figure 14 shows an example of a Lissajous curve scanned in this fashion. In practice, we use a much denser scan, but have shown this one to make the curve visible.

Fig. 14. Lissajous curve scanned in the focal plane of the collimating lens

Illumination lens. We use a 200 mm Nikon prime lens to collimate light from the above 'source' for its superior performance over off-the-shelf AR-coated achromatic doublets in terms of spherical and chromatic aberration, improving light efficiency and collimation. The output of the scene is cropped to a 1 inch-diameter circular beam and is passed through the beamsplitter cube apertures.

Interreflections. Interreflections are especially problematic in the case of temporally coherent light because they introduce strong spurious fringes. For example, light paths passing through a beamsplitter, and those reflected twice inside the beamsplitter also passing through interfere and cause strong fringes. Such fringes in the scene-only and reference-only images essentially nullify our contrast. We use optics with anti-reflective coatings designed for our laser wavelengths to reduce interreflections. We also found that deliberately misaligning optics by a small amount (sub-degree) helps avoid interreflections.

Beamsplitter. We use a thin 50:50 plate beamsplitter, since pellicle and cube beamsplitters cause strong interreflections. As above, we deliberately misalign optics to avoid interreflection artifacts.

Mirrors. We use high-quality mirrors of guaranteed $\lambda/4$ flatness to ensure a uniform phase reference throughout the field of view of the camera.

Translation stage. We use a translation stage from Newport with an accuracy of upto 10 nm and low-noise operation. For high-resolution depth recovery, it is important that the mirror positions images are captured at be accurate.

Camera lens. Our scenes are sized at the order of 1 inch. Therefore, we benefit from a lens that achieves high magnifications (1:1). This also allows for better contrast due to lower averaging of speckle (interference signal is convolved with the pixel box when captured with the camera). We use a 180 mm Canon prime macro lens in front of the camera.

Camera. We use a machine vision camera from AVT with a high sensitivity CCD sensor of resolution 8 MP, pixel size 3.5 µm, and a pixel pitch of 4 µm. Small pixel size increases interference contrast because we average interference speckle over a smaller area. In addition, it is important that the protective glass above the sensor be removed, since interreflections in the protective glass introduce strong fringes and nullify true interference contrast.

Neutral density filters. We use absorptive neutral density filters to make the intensities of both arms of the interferometer equal. Matching the brightness leads to an acquisition with optimal interference contrast.

Spectral filter. Ambient light reduces interference contrast because it adds to the intensity measurement at the camera but not to the interference. To reject ambient light, we use an ultra-narrowband spectral filter with a central wavelength around the wavelength of our lasers.

Alignment. Due to the long paths that light takes in the setup and the small (1 inch) aperture, the optical setup requires very careful alignment. Therefore, all of the setup is built around a cage system that reduces alignment requirements. The steering mirrors are aligned to make sure the beam at its mean position passes through the center of the beamsplitter box mounts. The reference mirror and camera are then aligned using the alignment technique described by Gkioulekas et al. [2015].

Component list. For easy reproducibility of the setup, we provide in Table 2 a list of the key components used in our implementation. We do not list standard parts used for mounting and positioning commonly available in optical labs.

D 3D RENDERINGS OF MEASURED DEPTHS

In Figure 15 and Figure 16, we render the depths recovered with our technique in Figures 6 and 7 of the main paper respectively.

E RECONSTRUCTION CODE

We provide basic Matlab code for recovering depth from measurements made with our method in Figure 17. The code assumes that the measurements are stored in a variable `frames` of size `H×W×4×4`, where `H` and `W` are the height and width of the measured images respectively, with the third dimension varying over sub-wavelength shifts and the fourth varying over four-bucket positions. The variable `scene` stores an ambient light image of the scene to serve as the guide image for the bilateral filter, and the variable `lam` denotes the synthetic wavelength. The function `bilateralFilter` executes bilateral filtering of its first argument with its second argument as the guide image with `spatialWindow` and `intensityWindow`.

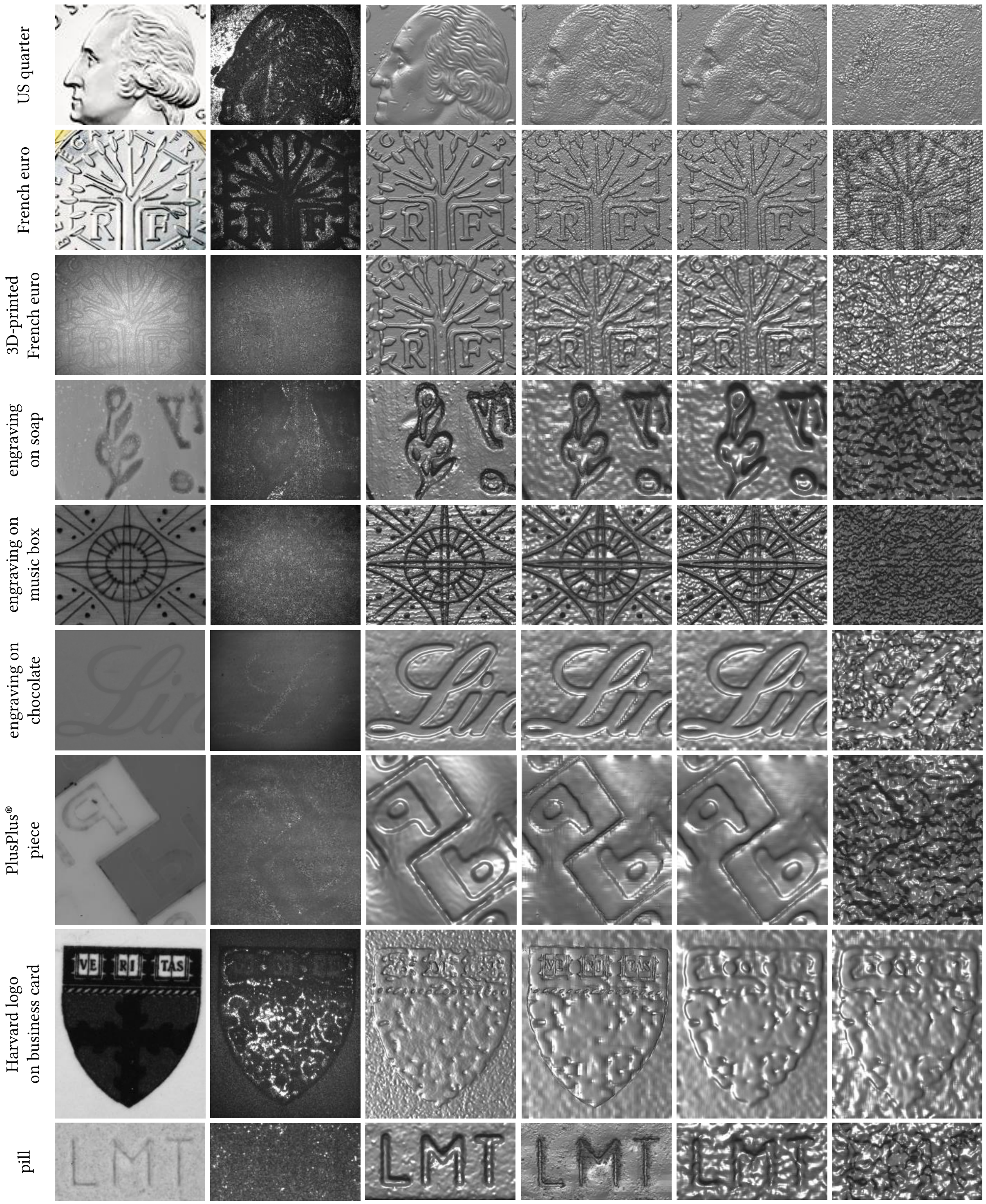


Fig. 15. Recovered depths for scenes with microscopic depth ranges rendered as surfaces. From left to right: ambient light image, one of our raw images, depth measured using OCT, our depth with bilateral filtering and direct-only probing, our depth with Gaussian filtering and direct-only probing, our depth with Gaussian filtering and no probing.

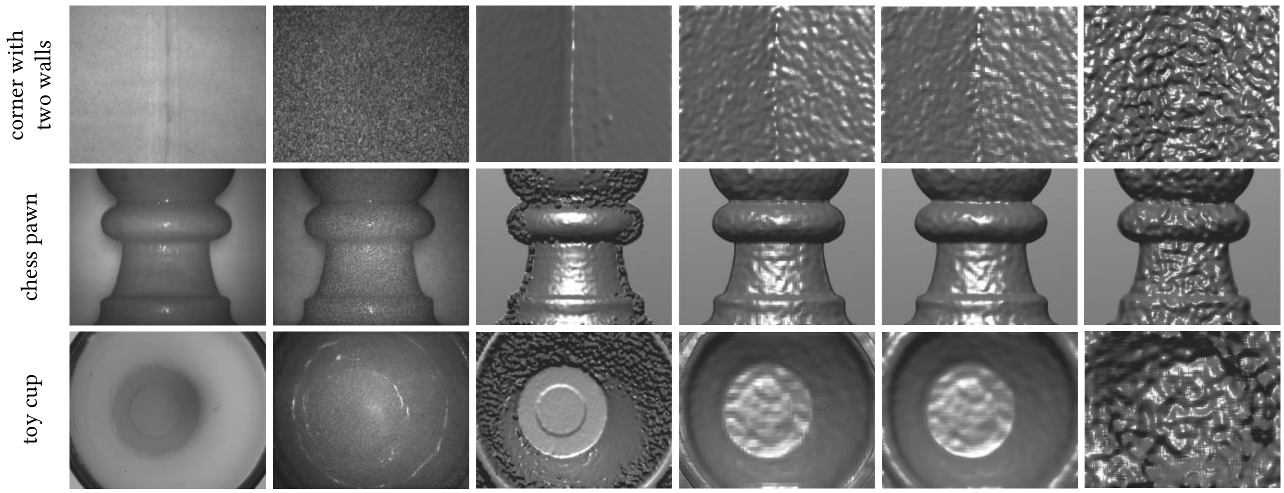


Fig. 16. Depth reconstruction results for scenes with macroscopic depth ranges rendered as surfaces. From left to right: ambient light image, one of our raw images, depth measured using OCT, our depth with bilateral filtering and direct-only probing, our depth with Gaussian filtering and direct-only probing, our depth with Gaussian filtering and no probing.

```
function depth = reconstruct(frames, lam, spatialWindow, ...
                             intensityWindow, scene)
  % Reconstruct depth from synthetic wavelength interferometry
  % frames:          HxWx4x4 array of measurements, where the third dimension
  %                  varies over subwavelength shifts and fourth over
  %                  four-bucket positions
  % lam:             synthetic wavelength
  % spatialWindow:   spatial window for the bilateral filter
  % intensityWindow: intensity window for the bilateral filter
  % scene:           ambient light image of the scene

  frames = im2double(frames)*4;
  scene = im2double(scene)*4;

  % Get interference-free images at each four-bucket position by averaging
  % images captured with sub-wavelength shifts
  meanFrames = mean(frames, 3);

  % Get correlation images at each four-bucket position by subtracting
  % interference-free images from the full images.
  correlation = frames - meanFrames;

  % Get the values of the correlation envelope by squaring and adding
  % correlation images
  envelope = squeeze(sum(correlation.^2, 3));

  % Filter the measured envelope with bilateral filtering using an ambient
  % light image of the scene as the guide image
  for position = 1:4
    envelope(:, :, position) = bilateralFilter(envelope(:, :, position), ...
                                               scene, spatialWindow, ...
                                               intensityWindow);
  end

  % Apply the four-bucket phase retrieval algorithm to estimate phase
  phase = atan2(envelope(:, :, 4) - envelope(:, :, 2), ...
                envelope(:, :, 1) - envelope(:, :, 3));

  % Convert phase to depth
  depth = phase*lam/(2*pi);

end
```

Fig. 17. Matlab code for recovering depth from our measurements