Macroscopic Interferometry: Rethinking Depth Estimation with Frequency-Domain Time-of-Flight

Achuta Kadambi, Jamie Schiel, and Ramesh Raskar

MIT Media Lab

{achoo,schiel,raskar}@mit.edu

Abstract

A form of meter-scale, macroscopic interferometry is proposed using conventional time-of-flight (ToF) sensors. Today, ToF sensors use phase-based sampling, where the phase delay between emitted and received, high-frequency signals encodes distance. This paper examines an alternative ToF architecture, inspired by micron-scale, microscopic interferometry, that relies only on frequency sampling: we refer to our proposed macroscopic technique as Frequency-Domain Time of Flight (FD-ToF). The proposed architecture offers several benefits over existing phase ToF systems, such as robustness to phase wrapping and implicit resolution of multi-path interference, all while capturing the same number of subframes. A prototype camera is constructed to demonstrate macroscopic interferometry at meter scale.

1. Introduction

Three-dimensional (3D) cameras capture the depth of objects over a spatial field. The recent emergence of low-cost, full-frame 3D cameras, such as Microsoft's Kinect, have facilitated many new applications in diverse areas of computer vision and graphics. These advances continue to drive demands for faster, more accurate, and more information-rich 3D systems that can operate outside of controlled environments.

Time of flight (ToF) cameras contain an active light-source that strobes coded patterns into a scene. The optical signal returning to the sensor exhibits a shift in phase corresponding to the propagation distance of the signal, which allows object depth to be calculated. This pervasive architecture, used in devices such as the Microsoft Kinect and Google Tango, is referred to as "phase ToF".

All phase ToF cameras rely on phase-sensing to determine object depths. Accurate estimation of phase becomes challenging in environments with, for example, multi-path interference. Phase is also periodic, meaning that it will wrap beyond a certain distance and create ambiguities in true object depths. Many workarounds have been proposed, but phase ToF still remains sensitive to multi-path, wrapping at extreme distances, and low levels of signal-to-noise-ratio (SNR).

1.1. Contributions

Inspired by optical coherence tomography (OCT), this paper recasts depth estimation into a macroscopic architecture, dubbed Frequency-Domain ToF (FD-ToF), avoiding several limitations with existing ToF cameras:

- Long-range objects that would ordinarily phase-wrap, can be ranged with no additional processing.
- Multi-path interference can be easily filtered with standard methods, such as Fourier Transformations, without need for increased sampling.
- Ranging is possible in environments with extremely low SNR.
- Drawing from existing theories in OCT, this paper provides a theoretical bound for multi-path separation.

2. Related Work

The connections between early techniques in interferometry, radar, and ToF imaging are well-known. In the 1990's, development of the modern ToF camera was inspired by the Michelson interferometer (see [26] for a historical overview). However, ToF range imaging has diverged from interferometric techniques, with little cross-pollination of ideas over the last decade. While recent work has linked ToF cameras with various radar systems [39]. In comparison, this paper aims to connect Frequency-Domain OCT to a complementary depth sensing architecture, Frequency-Domain ToF.

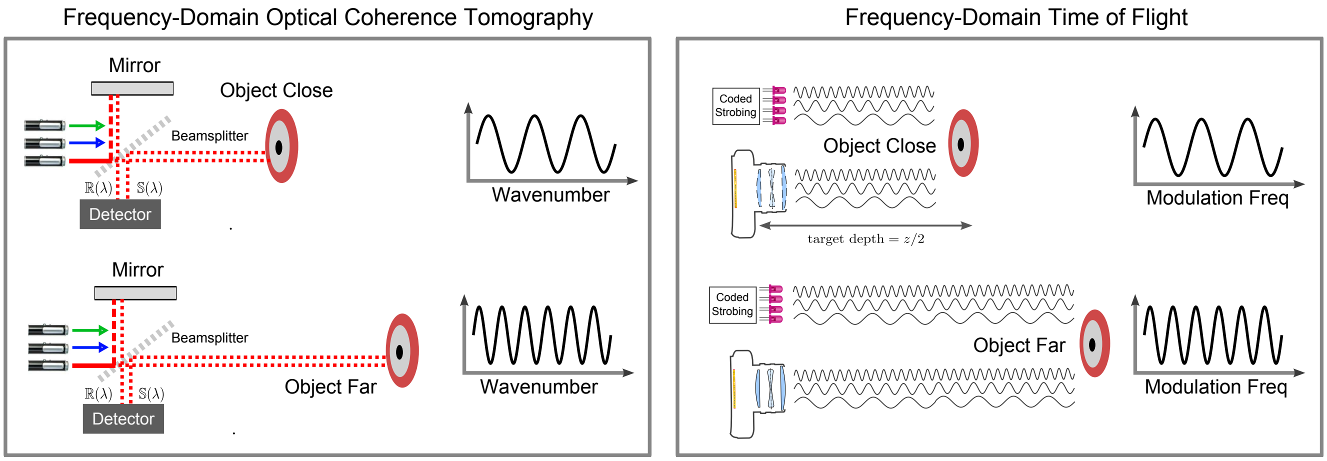


Figure 1. This paper repurposes the microscopic technique of Frequency-Domain OCT to a macroscopic technique, Frequency ToF. Both techniques encode optical time of flight in the frequency of the received waveform. For short optical paths (top row), the received signal in the primal-domain is lower in frequency than that of longer optical paths (bottom row).

Multifrequency time of flight describes a class of techniques where a *phase ToF* camera repeatedly samples the scene at multiple strobing frequencies (please note the emphasis). Our work is inspired by previous approaches in this line that leverage the rich phase-frequency information to recover a variety of light-transport metrics in complex scenes. Notable papers include the convex optimization algorithms by Heide et al. [13] and Lin et al. [25] as well as the spectral estimation methods of Bhandari et al. [3, 4] and Freedman et al. [6]. This paper is distinguished from prior art because it samples *only* in the frequency-domain, leading to a variant this paper refers to as “frequency ToF”. Since the underpinnings of frequency ToF are not new—having existed in the radar community for decades—this paper’s contribution lies in the empirical and theoretical evaluation of whether frequency ToF is a beneficial architecture for ToF cameras.

Disentangling multipath interference is one of the most popular topics in computational ToF imaging. The problem is posed as such: in phase ToF, reflections at different phase delays mix at a single pixel on the sensor. Extracting the constituent phases and amplitudes is a complicated signal processing problem, requiring additional measurements. We can classify prior approaches in the vein of multifrequency phase ToF [31, 22, 13, 3, 4, 6, 32, 5] or space-time light transport [34, 41, 42, 30, 40, 18, 10, 28, 19, 27, 21, 32].[1] The latter approach is complementary to this paper, as our approach is in the context of a single scene point and could be later extended with spatial analysis. Based on the intended goals, we consider the multifrequency phase ToF approaches of [31, 22, 3, 4] to be the closest related work and we use this as our point of comparison. If successful, a solution to multipath interference has scope to broader problems, such as ultrafast imaging [29, 38, 13, 20, 25, 15, 8, 9], scattered light imaging [37, 14], velocity imaging [12] and fluorescence imaging [2, 33, 1]. The list of references in mitigating multipath interference for ToF cameras is vast. Rather than proposing yet another specialized method, this paper introduces frequency ToF, where multipath correction is implicit in the Fourier Transform.

Phase wrapping is an artifact occuring in phase ToF where depth can only be recovered to a modulo factor. Phase unwrapping techniques address this problem by using multifrequency phase ToF, where each frequency yields phasor measurement [22, 3]. A Lissajous pattern is formed in phasor space by tracing out phasors at multiple frequencies. If the Lissajous pattern does not cross over in phasor space, the original scene depths can be uniquely recovered. However, for nearby points on the Lissajous pattern, the presence of shot noise can cause the Lissajous pattern to cross, leading to problems in unicity. Robust depth sensing at long-ranges has been achieved through the design of both an appropriate Lissajous pattern and incorporation of spatial or scene priors [11]. To avoid phase wrapping altogether, this paper explores Frequency-Domain ToF.

Optical coherence tomography is an optical interferometric technique used extensively in biomedical imaging [16]. Phase ToF was inspired by early, time-domain OCT systems, where the received signal is time-correlated with a reference signal to determine phase offset. Over the last decade, phase ToF literature has little to do with optical interferometry. This paper studies the relationship between the ToF camera sensor and the more recent interferometric technique of frequency-domain OCT. In frequency-domain OCT, 3D shape is obtained by illuminating a microscopic

[1]Time-coding doesn’t strictly fall into either category, but the few existing works are different from our approach [20, 36].

sample at multiple optical frequencies [7]. Transposing ideas from frequency-domain OCT to macroscopic 3D cameras offers a complementary way of thinking about ToF capture. To summarize: frequency ToF substitutes multiple optical frequencies for multiple temporal frequencies.

3. Preliminaries

A description of the basic principles of phase-based ToF is provided in Section 3.1, and a condensed overview of OCT in Section 3.2.

Terminology: The term *primal-domain* is used to refer to the original frame that a signal is sampled in. The term *dual-domain* refers to the frequency transform with respect to the primal-domain.

3.1. Phase ToF

A phase-based ToF camera measures the phase delay of optical paths to obtain depth by the relation

$$z = \frac{c\varphi}{2\pi f_{\mathrm{M}}} \quad d = z/2, \tag{1}$$

where z is the optical path length of reflection, d is the object depth, f_{M} is the modulation frequency of the camera, φ is the phase delay, and c is the speed of light. Modulation frequencies are typically around 20-50 MHz, corresponding to periods of 50-20 nanoseconds. To estimate φ with precision, a phase ToF camera uses an active illumination source strobed according to a periodic illumination signal. In standard implementations, the illumination strobing signal $i(t)$ can be modelled as a sinusoid

$$i\left(t\right) = \alpha \cos\left(2\pi f_{\mathrm{M}} t\right) + \beta_i, \tag{2}$$

where α is the modulation amplitude and β_i represents a DC component. At the sensor plane, the received optical signal $o(t)$ can be written as

$$o\left(t\right) = \Gamma\alpha \cos\left(2\pi f_{\mathrm{M}} t - \varphi\right) + \beta_o, \tag{3}$$

where $\Gamma \in [0, 1]$ is the attenuation in the reflected amplitude and now β_o includes DC offset from ambient light. Estimating the phase would require sampling Equation 3 multiple times. Such sampling in time-domain is challenging because it requires very short, nanosecond exposures. Instead, ToF cameras cross-correlate the optical signal with a reference signal at the same frequency, i.e., $r(t) = \cos\left(2\pi f_{\mathrm{M}} t\right)$. Without loss of generality set $\Gamma = 1$ to express the cross-correlation as

$$c(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} o(t) r(t+\tau) dt = \frac{\alpha}{2} \cos\left(2\pi f_{\mathrm{M}} \tau + \varphi\right). \tag{4}$$

Now, the primal-domain has changed from time, to τ. This plays a key role, as τ can be sampled at nanosecond timescales by varying a buffer between emitted and reference signals. To recover the phase and amplitude from the received signal, ToF cameras capture N subframes in the primal-domain of τ and, in software, compute an N-point Discrete Fourier transform (DFT) with respect to the primal. Suppose that four evenly spaced samples are obtained over the length of a period, for instance, $\tau = [0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}]^T$. Then the calculated phase can be written as

$$\varphi = \arctan\left(\frac{c(\tau_4) - c(\tau_2)}{c(\tau_1) - c(\tau_3)}\right), \tag{5}$$

and the calculated amplitude as

$$\alpha = \frac{1}{2}\sqrt{\left(c\left(\tau_4\right) - c\left(\tau_2\right)\right)^2 - \left(c\left(\tau_1\right) - c\left(\tau_3\right)\right)^2}. \tag{6}$$

The two real quantities of amplitude and phase can be compactly represented as a single complex number using phasor notation:

$$\mathbb{M} = \alpha e^{j\varphi}, \tag{7}$$

where $\mathbb{M} \in \mathcal{C}$ is the measured phasor, and j is the imaginary unit. Armed with the phase information, a ToF sensor computes depth using Equation 1 and provides a measure of confidence using the amplitude. This concludes our overview of the standard phase ToF operation.

Multi-path interference: Multi-path interference (MPI) occurs when K reflections return to the imaging sensor. The received signal can be written as

$$c_{\mathrm{MP}}(\tau) = \frac{1}{2}\left(\sum_{l=1}^{K} \alpha_l \cos\left(2\pi f_{\mathrm{M}} \tau + \varphi_l\right)\right) + \beta, \tag{8}$$

where the subscript MP denotes multi-path corrupted measurements. The received signal is now a summation of sinusoids at the same frequency but different phases. Obtaining the direct bounce, i.e., $K = 1$, is a very challenging problem. After simplification using Euler's identity, the measured amplitude and phase in the presence of interference can be written as

$$\varphi_{\mathrm{MP}} = \arctan\left(\frac{\sum_{i=1}^{K} \alpha_i \sin\varphi_i}{\sum_{i=1}^{K} \alpha_i \cos\varphi_i}\right) \tag{9}$$

$$\alpha_{\mathrm{MP}}^2 = \sum_{i=1}^{K} \alpha_i^2 + 2\underbrace{\sum_{i=1}^{K}\sum_{j=1}^{K} \alpha_i \alpha_j \cos\left(\varphi_i - \varphi_j\right)}_{i \neq j}. \tag{10}$$

In crux, phase ToF cameras encode multi-path interference as a summation of varying phases. Disentangling phases is a challenging, non-linear inverse problem that is well-known to be ill-conditioned at realistic levels of SNR [3, 6]. Later in this paper, multi-path is recast as a summation of varying frequencies, a solvable problem at low SNR.

Increasing the modulation frequency: Increasing the modulation frequency of phase ToF allows for greater depth precision [24]. Intuitively, at high frequencies, a small change in the estimated phase, corresponds to a small change in the estimated depth. The range accuracy ΔL is proportional to

$$\Delta L \propto c/f_{\mathrm{M}}. \tag{11}$$

Over the past few years, phase ToF hardware has supported increased modulation frequencies to boost depth precision. For instance, the new Kinect tripled the modulation frequency, increasing the modulation frequency to about 100 MHz. However, this frequency is at the upper limit of what the phase ToF architecture can handle. Without using phase unwrapping, scene objects at a depth greater than

$$d_{\text{ambiguity}} = c/2f_{\mathrm{M}}, \tag{12}$$

will encounter wrapping. Unwrapping, or disambiguation of phase, requires more measurements in combination with a lookup table and is susceptible to noise [23]. Another challenge at high modulation frequencies is an increase in the required sampling rate in the primal-domain. Frequency-Domain ToF does not have such drawbacks when using high strobing frequencies.

3.2. Primer on optical coherence tomography

Optical coherence tomography performs correlation directly on the optical signal using either: Time-Domain OCT (TD-OCT) to time-correlate an optical reference with a sample or Frequency-Domain OCT (FD-OCT) to sample only in frequency domain.[2] In this section, we provide a concise overview of FD-OCT, describing only the facets that can be applied to 3D cameras.

FD-OCT obtains depth by sampling the signal at different optical wavelengths (i.e. wavelength is the primal-domain). Figure 1 provides a schematic for the typical FD-OCT system. At a single wavelength, the detector receives an electric field from the reference object, which takes the form of

$$\mathbb{R}(\lambda) = \alpha(\lambda) e^{j\varphi_R(\lambda)}, \tag{13}$$

where $\mathbb{R}(\lambda)$ represents the received phasor as a function of optical wavelength. Similarly, the received electric field from the sample object is written as

$$\mathbb{S}(\lambda) = \alpha(\lambda) e^{j\varphi_S(\lambda)}. \tag{14}$$

Note that the amplitude of the sample and reference are assumed to be equal, which simplifies our explanation of the concept. A combination of the two reflections return to an imaging sensor. The electric field at the detector is the summation

$$\mathbb{M}(\lambda) = \frac{1}{2}\left(\mathbb{R}(\lambda) + \mathbb{S}(\lambda)\right), \tag{15}$$

where it is assumed that the constituent phasors are halved when they recombine (for instance, due to a beamsplitter). The current measured at the detector can be expressed as the real quantity

$$i(\lambda) = \frac{\eta q}{h\nu} \left|\mathbb{M}(\lambda)\right|^2, \tag{16}$$

where η is the detector sensitivity, q is the quantum of electric charge, h is Planck's constant, and ν is the optical frequency. By substituting Equation 15 into 16 we obtain

$$i(\lambda) = \tfrac{1}{4}\tfrac{\eta q}{h\nu}\left(\underbrace{\mathbb{R}(\lambda)\left(\mathbb{R}(\lambda)\right)^* + \mathbb{S}(\lambda)\left(\mathbb{S}(\lambda)\right)^*}_{\text{Autocorrelation}} + \underbrace{2Re\left(\left(\mathbb{R}(\lambda)\right)^* \mathbb{S}(\lambda) e^{-j\varphi_z}\right)}_{\text{Crosscorrelation}}\right). \tag{17}$$

Here, φ_z represents the phase delay due to the difference in optical path length between the reference and optical reflections. Similar to the ToF case, phase and z-distance are related:

$$\varphi_z = 2\pi z/\lambda. \tag{18}$$

In Equation 17 note that the Autocorrelation terms are DC with respect to the wavelength. By using this relation along with Equation 18, Equation 17 is rewritten as

$$i(\lambda) = \frac{1}{2}\frac{\eta q}{h\nu}\left(\alpha(\lambda)\right)^2\left(1 + 2\cos\left(\frac{2\pi z}{\lambda}\right)\right). \tag{19}$$

Now we introduce an auxiliary variable $k = 2\pi/\lambda$, which is known as the wave number. Equation 19 can be rewritten as

$$i(k) = \frac{1}{2}\frac{\eta q}{h\nu}\left(\alpha(k)\right)^2\left(1 + 2\cos(kz)\right), \tag{20}$$

where now the primal-domain is the wavenumber (k). The dual can be computed as

$$\mathcal{F}\left[i(k)\right](\kappa) \propto \delta(\kappa) + \delta(\kappa \pm z), \tag{21}$$

where κ represents the dual domain. To summarize: in Frequency-Domain OCT, reflections at multiple wavenumbers are measured at the detector and depth is encoded in the frequency of the received signal in primal-domain.

4. Frequency-Domain Time-of-Flight

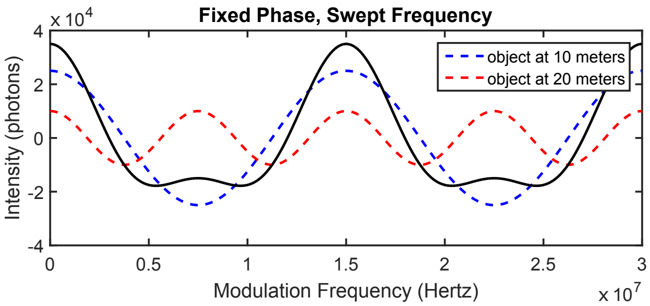
Inspired by Frequency-Domain OCT, the conventional operation of ToF cameras is re-examined. In this section a recipe for depth estimation is provided by sampling different modulation frequencies at a *single* phase step. This architecture is the proposed Frequency-Domain ToF, where the primal-domain is modulation frequency.

4.1. Depth sensing using only modulation frequency

Depth is calculated, not from phase steps, but from frequency steps. Recall from Section 3.1 that the received signal takes the form of

$$c(\tau) = \frac{\alpha}{2}\cos\left(2\pi f_{\mathrm{M}}\tau + \varphi\right) + \beta. \tag{22}$$

[2] Phase ToF was inspired by TD-OCT [26].



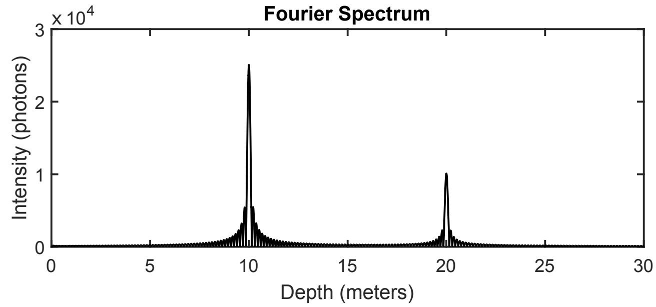


Figure 2. Multi-path interference is inherently resolved in Frequency-Domain ToF. (left) Multi-path interference from two reflections cause the measured signal (black) to have two tones (right). This simulation uses a frequency bandwidth of 1Ghz (left plot windowed for clarity).

In standard phase ToF, τ represents the primal-domain against which one would ordinarily compute the N-point DFT. Instead, consider substituting Equation 1 into Equation 4 to obtain

$$c(\tau, f_{\mathrm{M}}) = \frac{\alpha}{2} \cos\left(2\pi f_{\mathrm{M}} \tau + \frac{2\pi z}{c} f_{\mathrm{M}}\right) + \beta. \tag{23}$$

Without loss of generality assume that this signal is sampled only at the zero shift, i.e., $\tau = 0$. Then the received signal at the sensor takes the form of

$$c(\tau = 0, f_{\mathrm{M}}) = c(f_{\mathrm{M}}) = \frac{\alpha}{2} \cos\left(\frac{2\pi z}{c} f_{\mathrm{M}}\right) + \beta. \tag{24}$$

Now the primal domain is f_{M} and the associated dual takes the form of

$$\mathcal{F}\left[c\left(f_{\mathrm{M}}\right)\right](\kappa) \propto \delta(\kappa) + \delta\left(\kappa \pm \frac{2\pi z}{c}\right), \tag{25}$$

where κ is the dual-domain, corresponding to the inverse of the modulation frequency. Analogous to FD-OCT the depth can be obtained by finding the location of the support in the dual domain.

4.2. Multi-path interference in FD-ToF

An advantage of Frequency-Domain ToF is that multi-path interference is separable in the dual-domain. Recall that in the multi-path problem, K reflections return to the sensor and the received signal is given by

$$c(f_{\mathrm{M}}) = \frac{1}{2}\left(\sum_{l=1}^{K} \alpha_l \cos\left(\frac{2\pi z_l}{c} f_{\mathrm{M}}\right)\right) + \beta. \tag{26}$$

The associated Fourier transform may now be written as

$$\mathcal{F}\left[c\left(f_{\mathrm{M}}\right)\right](\kappa) \propto \delta(\kappa) + \sum_{l=1}^{K} \alpha_l \delta\left(\kappa \pm \frac{2\pi z_l}{c}\right). \tag{27}$$

Here, the multi-path corrupted signal is a sum of sinusoids at the same phase but at different frequencies. As shown in Figure 2, a Fourier transform allows resolution of multi-path interference.

4.3. Estimating frequency to obtain depth

Frequency-Domain ToF recasts depth estimation into the problem of frequency estimation of the received signal (Equation 27). The number of subframes required for depth estimation is the same as for phase ToF (since both rely on sinusoidal sampling). To estimate multi-path, the sampling rate must obey the Nyquist rate; therefore, no less than $2K$ samples can be taken to resolve K multi-path reflections. In theory, FD-ToF is susceptible to long-range artifacts because very large depths can alias to lower frequencies. This is avoided by appropriately choosing the sampling rate (for example, adhering to the Nyquist rate of the highest-frequency component, if it is known) in the primal-domain. The reader is directed to the supplement and Stoica et al. [35] for details about frequency estimation.

4.4. Frequency bandwidth and resolution

By casting the problem in the realm of OCT, recovery bounds can be provided for multi-path interference. Previous work in ToF literature [13, 20, 25] has not considered to what resolution multi-path reflections can be separated. The duality between OCT and Frequency-Domain ToF reveals the following:

Proposition 1: *Multi-path interference can be resolved in Frequency-Domain ToF if the optical path-length between any two reflections is less than:*

$$\Delta z \approx 0.6 c / \Delta f_{\mathrm{M}}. \tag{28}$$

Proof (Sketch): *The sampling function in the primal domain is a boxcar function* $\Pi\left(f_M\right) = H\left(f_M - f_M^-\right) - H\left(f_M - f_M^+\right)$*, where* f_M^- *and* f_M^+ *represent the minimum and maximum modulation frequencies that are sampled and* $H(\cdot)$ *refers to the Heaviside step function. The Fourier transform of* $\Pi\left(f_M\right)$ *takes the form of a scaled sinc function* $\mathcal{F}\left[\Pi\left(f_M\right)\right](\kappa) \propto \Delta f_M \frac{\sin \Delta f_M \kappa}{\Delta f_M \kappa}$*, where* $\Delta f_M = f_M^+ - f_M^-$*. The FWHM of this function determines the axial resolution* Δz*. After simplification (see supplement), this can be approximated as* $\Delta z \approx 0.6 c / \Delta f_M$*, the desired result.* $\square$

Not surprisingly, a larger frequency bandwidth Δf_{M} improves the optical path resolution.

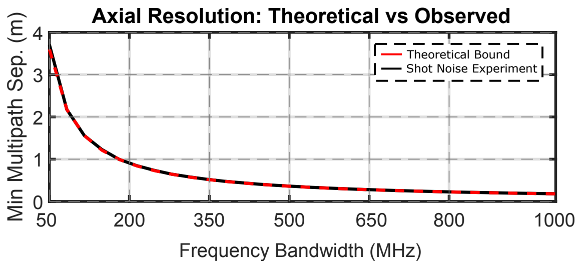


Figure 3. We show that the noiseless bound derived in *Proposition 1* is valid in typical shot noise limited scenarios.

5. Assessment and Results

Without any additions to hardware beyond the typical components of a ToF camera, and using fewer captured images, the proposed technique is shown to handle challenging scenes with multi-path interference, large distances, and low SNR. Comparisons were performed to a state-of-the-art paper that uses both multiple frequencies and phase steps to mitigate multi-path interference (hereafter Bhandari's method) [3].

5.1. Synthetic scenes

Resolution bound: *Proposition 1* describes an upper bound on multi-path resolution, derived in the main paper for the noiseless case. This bound is based on the sampling properties of the system (i.e. highest and lowest frequencies sampled) and does not factor noise. However, Figure 3 shows that at SNRs, governed by shot-noise, of a typical camera, the observed resolution very nearly approaches the best-case resolution bound. To our knowledge, this is one of only a few papers in ToF imaging to analyze the minimum separation needed to accurately resolve multi-path. The supplement contains a derivation of the resolution bound.

Dragon scene: A complex scene with transparencies and caustics was generated using the Mitsuba renderer [17]. Figure 4 displays the rendered scene, consisting of a dragon behind a transparent sheet. Two dominant reflections are observed, from the transparency and the dragon behind. The camera noise model includes read noise, dark noise, and shot noise. Three scenarios are compared across SNR: Frequency-Domain ToF (12 subframes), phase ToF (12 subframes), and Bhandari's method (using 120 subframes). As expected, Frequency-Domain ToF performs better than phase ToF at all levels of noise. At high levels of SNR, Bhandari's technique of disentangling phases is the best performer. However, under a shot-noise limited model, the measurement SNR is typically between 15 and 40 dB. Within this regime, our proposed technique outperforms Bhandari's method despite using far fewer subframes.

5.2. Experiments

Physical implementation: Physical experiments were conducted by reconfiguring a sensor used previously for

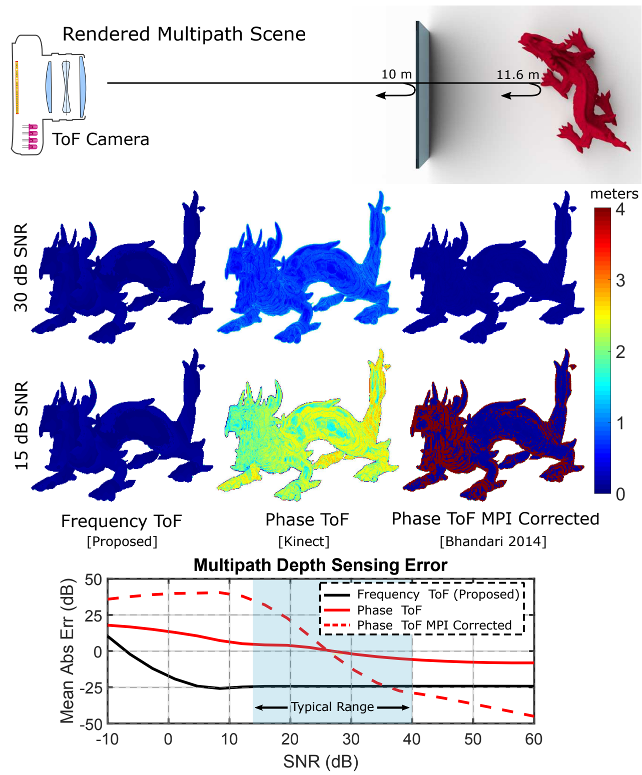


Figure 4. Frequency-Domain ToF is more robust than other methods at low levels of SNR. (Top) The rendered scene consists of a partially occluded dragon. (Middle) Registered plots of mean absolute error at (Upper Row) 15 and (Lower Row) 30 dB SNR. (Bottom) Plot of error across varying levels of SNR. Bhandari's method [3] is shown in the red dashed line.

coded ToF imaging [20]. Dynamic reconfiguration of modulation frequencies is achieved by controlling the sensors electronic exposure and light strobing with a phase-locked loop of an Altera Cyclone IV FPGA (Figure 5a). The CMOS sensor is a 120x160 pixel PMD 19k-S3 ToF sensor and the illumination bank is comprised of six, diffused, 650nm Mitsubishi LPC-836 laser-diodes. Calibration was performed at each modulation frequency to remove non-linear effects. All physical results in this paper are obtained by capturing either 4 or 45 frequency-indexed subframes within an achievable sensor bandwidth of 5 to 50MHz.

Window scene: An arrangement of boxes was placed behind a thick glass slab. Two dominant reflections return to the camera: one from the glass and another from the background scene. Using the existing method of phase ToF leads to a mean absolute error of 63 cm—note that the depth is underestimated due to the mixture from the foreground object. The proposed technique of FD-ToF recovers the background depth with a mean absolute error of 5cm. This is in line with previous approaches in multi-depth imaging [20, 3], but FD-ToF allows scene recovery using only four subframes, the same as phase ToF.

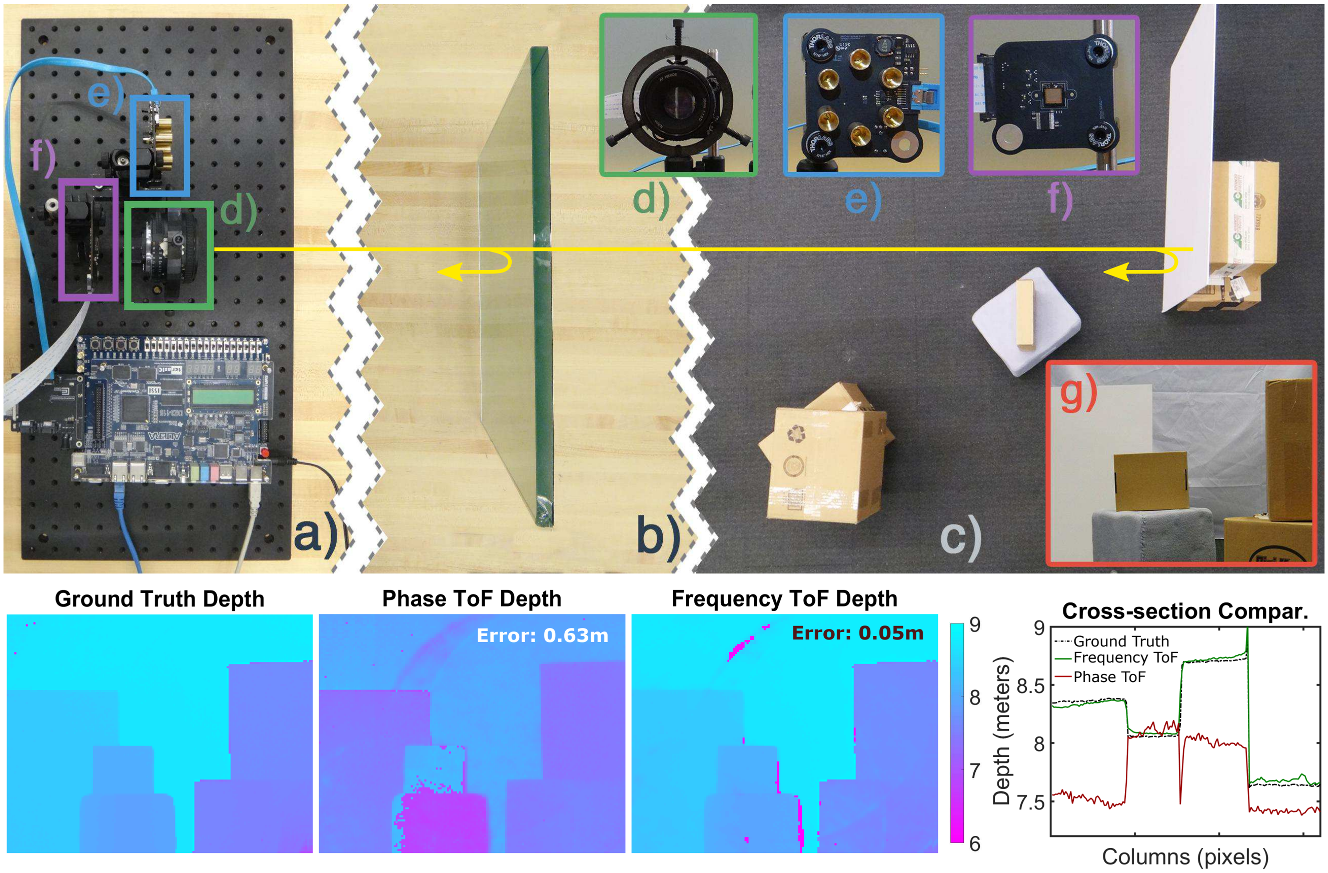


Figure 5. Window scene. (Top Figure) Overview of the prototype FD-TOF camera hardware and multi-path test scene. The camera (a) looks through a thick glass slab (b), which creates multi-path interference, at an arrangement of boxes (c). This is a medium-range scene; the wavy dash-lines indicate a large separation. The glass slab is 4.75m in front of the sensor plane, and the the first box is 7.4m in front. Our prototype camera with zoomed in lens (d), lasers (e), and bare sensor (f). The lower-right inset (g) shows the approximate field-of-view of the system. (Bottom Row) shows recovered depths, using the mean absolute error metric.

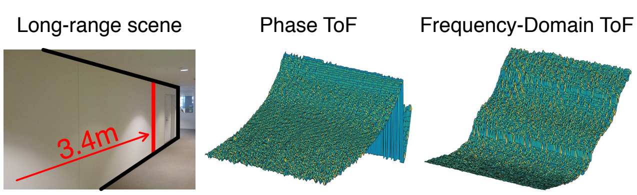


Figure 6. Phase is implicitly unwrapped in Frequency-Domain ToF. At 44 MHz the wrapping depth is 3.4m.

Long-range scene: Figure 6 shows a long-range capture of an angled wall. To isolate the effect of phase wrapping, no post-processing was performed on either FD-ToF or phase ToF (this includes radial calibration). Collected at a frequency of 44 MHz, phase ToF exhibits a wrapping artifact at 3.4 meters. In comparison, while the recovered depth from FD-ToF is noisy (due to limitations with the prototype hardware), wrapping artifacts are not present.

Everyday scene: Multi-path interference is far from a niche problem, occuring in everyday scenes. A living room scene is shown in Figure 7 with a chair and lamp shade, concave objects that will exhibit diffuse multi-path interference. Using 4 subframes, phase ToF over-estimates concave portions of the chair by about 20 centimeters. When sampling 4 frames, Frequency-Domain ToF recovers a depth map that is robust to multi-path, but exhibits noise. Using 45 subframes, FD-ToF recovers the scene with higher fidelity than Bhandari's method, which requires 180 subframes. Error is quantified in the upper-right of Figure 7.

6. Discussion and Limitations

In summary, this paper has described a complementary technique to obtain depth by sampling only in the frequency domain. Practical benefits are demonstrated, including accurate multi-path correction, immunity to phase wrapping and scene recovery at low SNR. In comparison to other multi-path correction schemes, the proposed technique reduces depth error, while requiring fewer subframes.

Diffuse and Specular Multi-path: Few papers in literature have proposed solutions that are robust to both diffuse

Method	Multi-path error [m]	Single-path error [m]	Overall error [m]
4-FD-ToF	0.319	0.044	0.187
45-FD-ToF	**0.153**	0.020	**0.090**
Phase ToF	0.347	**0.016**	0.191
Bhandari	0.279	0.018	0.145

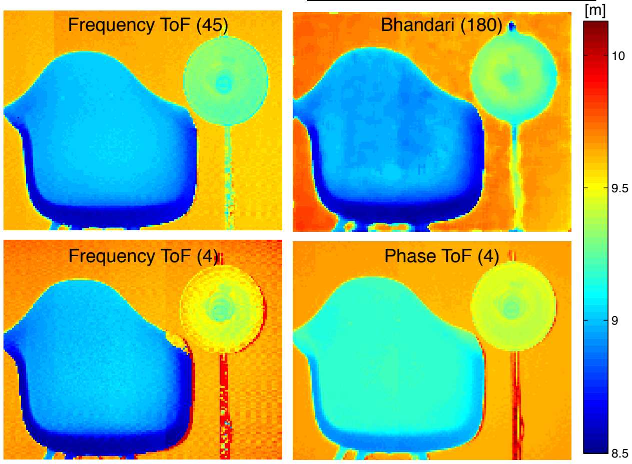


Figure 7. Naive phase ToF overestimates the chair's depth by 20cm due to diffuse multi-path. The proposed technique outperforms the state-of-the-art MPI correction scheme in [3], while using fewer subframe measurements (45 vs 180).

and specular multi-path. For example, Gupta et al. [10] and Naik et al. [28] assume smoothness in global illumination; their models work for diffuse multi-path but not specular multi-path. In contrast, Kadambi et al. [20] assume sparse multi-path; their model works for specular but not diffuse multi-path. By recasting depth estimation in frequency, both diffuse (as in Figure 7) and specular (as in Figure 5) multi-path can be addressed.

Simplified pipeline: In commercial depth sensors, the phase ToF pipeline can occur in stages:

Phase Subframes → DFT → Unwrap → MPI Correct

Even though some steps, such as phase wrapping, seem easy to solve, small errors can magnify as they propagate through the pipeline. In comparison, the proposed technique of FD-ToF exhibits a sparser pipeline:

Frequency Subframes → DFT

Number of subframes required: Previous methods for multi-path correction require many subframes, adopt restrictive light transport assumptions, or employ elaborate hardware schemes. Gupta et al. [10] and Naik et al. [28] achieve results with only 4 and 5 subframes by limiting scope to diffuse multi-path only. Some models are more general [3], but require dozens of subframes. Table 1 shows the *minimum* number of measurements that are required to handle multi-path interference; in the presence of noise or model mismatch, most solutions require many more subframes. In this paper, FD-ToF has demonstrated better results than [3], while using a quarter of the subframes.

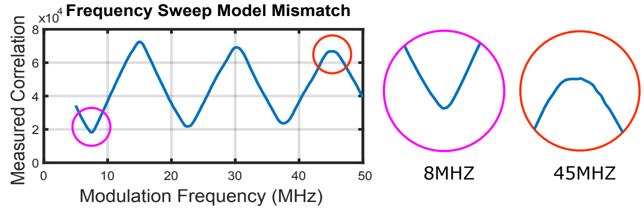


Figure 8. Signal nonlinearities cause ranging errors. At 8MHz the waveform resembles a triangle wave; at 45 MHz it resembles a sinusoid. This is a limitation of the prototype hardware.

Table 1. Minimum subframes req'd to handle multipath interference.

Paper	Frames
Bhandari et al. [3]	15
Freedman et al. [6]	9
Kirmani et al. [22]	8
Naik et al. [28]	5
Gupta et al. [10]	4
FD-ToF	**4**

Limitations of prototype: Since the technique is prototyped on a phase ToF CMOS sensor, the hardware is suited for phase but not frequency sweeps. This particular chip is designed to work over a narrow range of frequencies—outside this range, the signal is no longer sinusoidal (Figure 8). In addition, the sensor is geared to rapidly sweep in phase not frequency, precluding us from showing real-time results. Because the proposed method has been shown to work with fewer subframes than previous techniques, there is no fundamental limitation to achieving real-time performance.

Limitations of method: Following *Proposition 1*, resolving multi-path returns is limited to about 3 meters depth with the current bandwidth of 45 MHz. Until the bandwidth increases, FD-ToF is perhaps less suited for entertainment applications and better suited for longer-range depth sensing. In addition, although FD-ToF is robust to phase wrapping, long range depths can cause aliasing. However, as shown in the supplement, such aliasing will only occur in impossibly long-range scenarios.

Conclusion: Rather than directly solving challenges with existing phase ToF sensors, this paper describes a practical method to realize macroscopic interferometry on a standard ToF sensor.

Acknowledgments: The authors thank Vage Taamazyan, Suren Jayasuriya and the anonymous reviewers for valuable feedback. Achuta Kadambi is supported by the Charles Draper Doctoral Fellowship and the Qualcomm Innovation Fellowship.

References

[1] A. Bhandari, C. Barsi, and R. Raskar. Blind and reference-free fluorescence lifetime estimation via consumer time-of-flight sensors. *Optica*, 2(11):965–973, 2015. 2

[2] A. Bhandari, C. Barsi, R. Whyte, A. Kadambi, A. J. Das, A. Dorrington, and R. Raskar. Coded time-of-flight imaging for calibration free fluorescence lifetime estimation. In *OSA Imaging Systems and Applications*, 2014. 2

[3] A. Bhandari, M. Feigin, S. Izadi, C. Rhemann, M. Schmidt, and R. Raskar. Resolving multipath interference in kinect: An inverse problem approach. In *IEEE Sensors*, 2014. 2, 3, 6, 8

[4] A. Bhandari, A. Kadambi, and R. Raskar. Sparse linear operator identification without sparse regularization? applications to mixed pixel problem in time-of-flight/range imaging. In *ICASSP*, 2014. 2

[5] M. Feigin, R. Whyte, A. Bhandari, A. Dorington, and R. Raskar. Modeling wiggling as a multi-path interference problem in amcw tof imaging. *OSA Optics Express*, 2015. 2

[6] D. Freedman, Y. Smolin, E. Krupka, I. Leichter, and M. Schmidt. Sra: Fast removal of general multipath for tof sensors. In *ECCV*. 2014. 2, 3, 8

[7] J. Fujimoto. Optical coherence tomography. In *OSA Conference on Lasers and Electro-Optics*, 2006. 3

[8] G. Gariepy, N. Krstajic, R. Henderson, C. Li, R. R. Thomson, G. S. Buller, B. Heshmat, R. Raskar, J. Leach, and D. Faccio. Single-photon sensitive light-in-flight imaging. 2

[9] I. Gkioulekas, A. Levin, F. Durand, and T. Zickler. Micron-scale light transport decomposition using interferometry. *ACM Transactions on Graphics*, 2015. 2

[10] M. Gupta, S. K. Nayar, M. B. Hullin, and J. Martin. Phasor imaging: A generalization of correlation-based time-of-flight imaging. 2014. 2, 8

[11] M. Hansard, S. Lee, O. Choi, and R. Horaud. Disambiguation of time-of-flight data. In *Time-of-Flight Cameras*, pages 29–43. Springer, 2013. 2

[12] F. Heide, W. Heidrich, M. Hullin, and G. Wetzstein. Doppler time-of-flight imaging. *ACM Transactions on Graphics*, 2015. 2

[13] F. Heide, M. B. Hullin, J. Gregson, and W. Heidrich. Low-budget transient imaging using photonic mixer devices. *ACM Transactions on Graphics*, 2013. 2, 5

[14] F. Heide, L. Xiao, A. Kolb, M. B. Hullin, and W. Heidrich. Imaging in scattering media using correlation image sensors and sparse convolutional coding. *Optics express*, 2014. 2

[15] B. Heshmat, G. Satat, C. Barsi, and R. Raskar. Single-shot ultrafast imaging using parallax-free alignment with a tilted lenslet array. In *CLEO: Science and Innovations*, pages STu3E–7. Optical Society of America, 2014. 2

[16] D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M. R. Hee, T. Flotte, K. Gregory, C. A. Puliafito, et al. Optical coherence tomography. *Science*, 1991. 2

[17] W. Jakob. Mitsuba renderer, 2010. 6

[18] A. Jarabo, J. Marco, A. Munoz, R. Buisan, W. Jarosz, and D. Gutierrez. A framework for transient rendering. *ACM Transactions on Graphics*, 2014. 2

[19] S. Jayasuriya, A. Pediredla, S. Sivaramakrishnan, A. Molnar, and A. Veeraraghavan. Depth fields: Extending light field techniques to time-of-flight imaging. *3DV*, 2015. 2

[20] A. Kadambi, R. Whyte, A. Bhandari, L. Streeter, C. Barsi, A. Dorrington, and R. Raskar. Coded time of flight cameras: sparse deconvolution to address multipath interference and recover time profiles. *ACM Transactions on Graphics*, 2013. 2, 5, 6, 8

[21] A. Kadambi, H. Zhao, B. Shi, and R. Ramesh. Occluded imaging with time of flight sensors. *ACM Transactions on Graphics*, 2015. 2

[22] A. Kirmani, A. Benedetti, P. Chou, et al. Spumic: Simultaneous phase unwrapping and multipath interference cancellation in time-of-flight cameras using spectral methods. In *IEEE ICME*, 2013. 2, 8

[23] A. Kolb, E. Barth, R. Koch, and R. Larsen. Time-of-flight sensors in computer graphics. In *Proc. Eurographics (State-of-the-Art Report)*, 2009. 4

[24] R. Lange and P. Seitz. Solid-state time-of-flight range camera. *Quantum Electronics, IEEE Journal of*, 2001. 4

[25] J. Lin, Y. Liu, M. B. Hullin, and Q. Dai. Fourier analysis on transient imaging with a multifrequency time-of-flight camera. 2, 5

[26] X. Luan. Experimental investigation of photonic mixer device and development of tof 3d ranging systems based on pmd technology. *Diss., Department of Electrical Engineering and Computer Science, University of Siegen*, 2001. 1, 4

[27] N. Matsuda, O. Cossairt, and M. Gupta. Mc3d: Motion contrast 3d scanning. In *Computational Photography (ICCP), 2015 IEEE International Conference on*, pages 1–10. IEEE, 2015. 2

[28] N. Naik, A. Kadambi, C. Rhemann, S. Izadi, R. Raskar, and S. B. Kang. A light transport model for mitigating multipath interference in time-of-flight sensors. 2, 8

[29] N. Naik, S. Zhao, A. Velten, R. Raskar, and K. Bala. Single view reflectance capture using multiplexed scattering and time-of-flight imaging. In *ACM Transactions on Graphics (TOG)*, volume 30, page 171. ACM, 2011. 2

[30] M. O'Toole, F. Heide, L. Xiao, M. B. Hullin, W. Heidrich, and K. N. Kutulakos. Temporal frequency probing for 5d transient analysis of global light transport. *ACM Transactions on Graphics*, 2014. 2

[31] A. D. Payne, A. P. Jongenelen, A. A. Dorrington, M. J. Cree, and D. A. Carnegie. Multiple frequency range imaging to remove measurement ambiguity. 2009. 2

[32] C. Peters, J. Klein, M. B. Hullin, and R. Klein. Solving trigonometric moment problems for fast transient imaging. *ACM Transactions on Graphics*. 2

[33] G. Satat, B. Heshmat, C. Barsi, D. Raviv, O. Chen, M. G. Bawendi, and R. Raskar. Locating and classifying fluorescent tags behind turbid layers using time-resolved inversion. *Nature communications*, 2015. 2

[34] S. M. Seitz, Y. Matsushita, and K. N. Kutulakos. A theory of inverse light transport. In *IEEE ICCV*, 2005. 2

[35] P. Stoica and R. L. Moses. *Introduction to spectral analysis*, volume 1. Prentice hall Upper Saddle River, NJ, 1997. 5

[36] R. Tadano, A. Pediredla, and A. Veeraraghavan. Depth selective camera: A direct, on-chip, programmable technique for depth selectivity in photography. In *IEEE ICCV*, 2015. 2

[37] A. Velten, T. Willwacher, O. Gupta, A. Veeraraghavan, M. G. Bawendi, and R. Raskar. Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging. *Nature Communications*, 2012. 2

[38] A. Velten, D. Wu, A. Jarabo, B. Masia, C. Barsi, C. Joshi, E. Lawson, M. Bawendi, D. Gutierrez, and R. Raskar. Femto-photography: Capturing and visualizing the propagation of light. *ACM Transactions on Graphics*, 2013. 2

[39] R. Whyte, L. Streeter, M. Cree, and A. Dorrington. Application of lidar techniques to time-of-flight range imaging. *Applied Optics*, 2015 (To Appear). 1

[40] R. Whyte, L. Streeter, M. J. Cree, and A. A. Dorrington. Resolving multiple propagation paths in time of flight range cameras using direct and global separation methods. *Optical Engineering*, 54(11):113109–113109, 2015. 2

[41] D. Wu, A. Velten, M. OToole, B. Masia, A. Agrawal, Q. Dai, and R. Raskar. Decomposing global light transport using time of flight imaging. *International Journal of Computer Vision*, 2014. 2

[42] D. Wu, G. Wetzstein, C. Barsi, T. Willwacher, Q. Dai, and R. Raskar. Ultra-fast lensless computational imaging through 5d frequency analysis of time-resolved light transport. *International Journal of Computer Vision*, 110(2):128–140, 2014. 2