

Received October 9, 2017, accepted November 15, 2017, date of publication November 17, 2017, date of current version December 22, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2775138

Rethinking Machine Vision Time of Flight With GHz Heterodyning

ACHUTA KADAMBI¹, (Student Member, IEEE), AND RAMESH RASKAR

Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Corresponding author: Achuta Kadambi (e-mail: achoo@mit.edu)

This work was supported by the Media Laboratory consortium funding. The work of A. Kadambi was supported in part by the Qualcomm Innovation Fellowship and in part by the Draper Doctoral Fellowship.

ABSTRACT Time-of-flight (ToF) 3-D cameras like the Microsoft Kinect, are prevalent in computer vision and computer graphics. In such devices, the power of an integrated laser is amplitude modulated at megahertz frequencies and demodulated using a specialized imaging sensor to obtain subcentimeter range precision. To use a similar architecture and obtain micrometer range precision, this paper incorporates beat notes. To bring telecommunications ideas to correlation ToF imaging, we study a form of “cascaded Time of Flight” which uses a hertz-scale intermediate frequency to encode high-frequency pathlength information. We show synthetically and experimentally that a bulk implementation of opto-electronic mixers offers: 1) robustness to environmental vibrations; 2) programmability; and 3) stability in frequency tones. A fiberoptic prototype is constructed, which demonstrates 3- μm range precision over a range of 2 m. A key contribution of this paper is to study and evaluate the proposed architecture for use in machine vision.

INDEX TERMS Machine vision, computer graphics, imaging.

I. INTRODUCTION

Time of flight (ToF) imagers are a popular method of obtaining meter-scale depth images of a scene. Their acceptance in machine vision stems from a few unique advantages; for example, the ability to obtain depth from scenes with a variety of textures and materials or the capability to sense depth in a monocular fashion.

Centric to all ToF imagers is the relation

$$\text{Distance} = \text{Velocity} \times \text{ToF}. \quad (1)$$

Since the velocity of light is known ($c = 3 \times 10^8$ meters per second), a measurement of the ToF would allow distance estimation from Equation 1. Unfortunately, measuring the ToF with the required degree of precision is a challenging task. For a more complete picture, the reader is referred to review articles on time of flight imaging [1]–[3].

The core culprit—one that makes ToF imager design difficult—is the speed of light. Consider that in one billionth of a second, a photon travels a distance of one imperial foot. An imaging system with a temporal precision of a nanosecond would therefore have a depth precision of an imperial foot—unacceptable for computer vision and graphics applications. To obtain centimeter and sub-cm resolution, ToF imagers need to push the frontiers of sensing, to operate at speeds faster than a billionth of a second.

In this paper, we focus our study on correlation ToF imagers, a robust method to estimate the time of flight used in many robotic and automotive applications [3], [4]. A correlation ToF imager illuminates the scene with a strobed light source that modulates its intensity in a high-frequency, sinusoidal pattern. When light reflects back to the camera a phase shift is observed in the received intensity signal. This phase shift is directly proportional to both the distance traveled and the chosen frequency of strobed illumination.

Higher frequencies are known to improve the depth resolution of correlation ToF. As such, there has been a continued quest by hardware engineers to build high-frequency correlation ToF imagers. Figure 1 illustrates a timeline of a few, selected commodity ToF imagers. At the time of this writing, the commodity ToF imager with the highest achievable frequency is the Microsoft Kinect which operates at 120 MHz [5]. Unfortunately, silicon advances in image sensors are not likely to result in the ability to detect GHz frequencies anytime soon (henceforth, the “GHz gap”). Therefore, our aim in this paper is to incorporate heterodyning into the correlation ToF pipeline, to achieve high depth resolution. We are concerned with very low heterodyne frequencies, on the order of a few Hertz, to motivate future work where a video framerate sensor could be used for measurements.

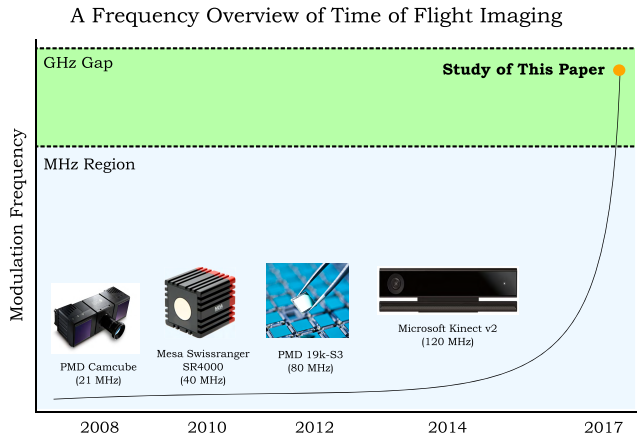


FIGURE 1. The GHz gap in correlation Time of Flight imaging. The modulation frequencies of correlation ToF imagers have steadily increased over the past decade. The Microsoft Kinect at 120 MHz has the highest modulation frequency among commodity devices. This paper extends the scope of modulated ToF imaging to the realm of GigaHertz frequencies (green region).

In recent years, the optics community has presented many solutions for compressing high-frequency range information into low-frequency, detectable signals using beat frequencies. A closely related method is “superheterodyne interferometry”, where multiple laser sources are used to create synthetic wavelengths. The technique dates back to a seminal paper by Dandliker *et al.* [6], and continues to attract interest in follow-up work [7], [8]. Most recently, superheterodyne interferometry has been extended to 1550nm wavelengths in an interesting fiberoptic approach by Li *et al.* [9]. Despite the promise of superheterodyne interferometry, there are a few limitations which limit its use as a commodity system. Specific limitations with superheterodyne implementations include:

- (A) Vibrations affecting bulk optical paths;
- (B) Frequency drift between multiple lasers; and
- (C) Partial tunability in synthetic wavelength.

In this paper, we show that a bulk fiberoptic setup of cascaded modulation can be used to achieve similar results to the the 2017 paper by Li *et al.* [9], but with improved performance in context of (A-C).

A. SCOPE AND CONTRIBUTIONS

Heterodyning is an elementary principle in a range of communities (radar, telecom., optics, etc). Our primary contribution is to study the heterodyne principle in the novel context of correlation ToF imagers (e.g. as used in computer vision). We first devise a cascaded modulation scheme that uses correlation sampling to parallel the sensing architecture of the Microsoft Kinect. Subsequently, we generalize the approach to a cascade of K elements, to enable frequency boosting and study the behavior from the Jacobi-Anger approximation. In describing this architecture, we find that the limitations of contrasting techniques (e.g., bullets A-C above) can be overcome with a bulk fiberoptic setup. We then build this setup

to empirically study the tolerance of the cascaded method to vibrations, frequency stability, and tunability in the beat note. The experimental results demonstrate 10 micron range precision from a Hertz-level beat note. We believe this shows the promise of using a video-framerate detector to collect time of flight measurements in harsh 3D vision conditions. Finally, we formulate computer graphics path-traced renderings to show the implications of scaling this technique to full-field 3D. Our contributions, when taken together, suggest promise in using beating for correlation ToF imaging.

II. RELATED WORK

An inspiration for writing this paper stems from research in computational imaging that describes the benefits of having higher ToF modulation frequencies. In 2001, Lange and Seitz [10] devised an error bound on the variance of range estimation based upon the modulation frequency (and other factors). Since, then various authors have used a sweep of multiple modulation frequencies to address multipath interference [11]–[16]. In particular, Kadambi *et al.* [16] derive a theoretical bound showing that multipath separation is easier at high frequency bandwidths. In earlier work, Gupta *et al.* uses not a sweep, but a single, high frequency of modulation to demonstrate that global illumination vanishes at high frequencies [17]. A frequency analysis of time of flight imaging is also applicable to unconventional applications. Heide *et al.* [18] and Lin *et al.* [19] use a dense assortment of frequencies to recover transient images of a scene at lower cost than existing solutions (e.g. streak imaging [20]). Building upon work from [21], [22], and Kadambi *et al.* theoretically analyze the impact of modulation frequency on spatial recovery of scenes around a corner, concluding that higher frequencies are important to see around corners [23]. All of the above approaches were conducted in the context of homodyne imaging (where the light and camera modulate and demodulate at the same frequency). Recent work has analyzed ToF imaging in the context of heterodyne approaches, to various ends [24]–[27].

A variety of methods exist for high-frequency ToF imaging (above MHz frequencies), such as chirped modulation, direct detection, and heterodyning with an optical local oscillator [28], [29]. These methods are not designed to obtain micron resolution using a Hertz-level beat note. For example, optical interferometry analyzes the phase shift between light wavefronts and can therefore be seen as a THz frequency ToF imager, allowing nanometer scale pathlength resolution [30], [31]. While the pathlength resolution is phenomenal, interferometry is limited in its practicality: (1) the frequency of light is so high that environmental disturbances can cause nuisance phase shifts; and (2) the measurement signal diminishes in the presence of vibrations. Two incoherent approaches for high-frequency imaging are SPAD and streak cameras—both have been used in computer vision and graphics (cf. streak techniques in [20], [21], [32], and [33] and SPAD techniques in [34]–[37]). However, both methods require sophisticated optical equipment and the pathlength

resolution for both devices is the order of a millimeter. The cascaded method described in this paper demonstrates micrometer range precision.

III. CORRELATION TIME OF FLIGHT IMAGING

A description of the basic principles of correlation ToF imaging is provided in Section III-A, and the impact of modulation frequency is discussed in Section III-B.

A. PRINCIPLE OF RANGE ESTIMATION

Correlation ToF is an active imaging system. A light source, like an LED or laser, strobes in a sinusoidal pattern, such that the optical power is written as

$$p(t) = \alpha \cos(\Omega t) + \beta_i, \quad (2)$$

where $p(t)$ is the time-varying optical power, α is the modulation amplitude and β_i is the DC component of modulation, where $\beta_i \geq \alpha$ (since light cannot have negative power). The camera receives a time-delayed version of the outgoing power, written as

$$p(t - \tau) = \Gamma \alpha \cos(\Omega t - \underbrace{\Omega \tau}_{\varphi}) + \beta_o \quad (3)$$

where τ is the time of flight, $\Gamma \in [0, 1]$ is the attenuation in the reflected amplitude and β_o is the DC component which now includes ambient light (hence, $\beta_o \geq \beta_i$). Since τ is on the order of nanoseconds, only a large value of Ω will cause an appreciable phase shift (in current systems, $\Omega \approx 10^8$ Hz). A correlation ToF camera estimates φ through correlation of the received signal with a reference signal. Define the reference signal as

$$r(t) = \cos(\Omega t), \quad (4)$$

noting that the reference signal can have negative values. Without loss of generality set $\Gamma = 1$, and assume a quantum efficiency of unity to write the correlation as

$$\begin{aligned} c(m) &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_{T/2}^{T/2} p(t - \tau) r(t + m) dt \\ &= \frac{\alpha}{2} \cos(\Omega m + \Omega \tau). \end{aligned} \quad (5)$$

Since distance is the product of velocity and time (cf Equation 1), one can relate the time of flight to the depth as

$$\tau = \frac{z_{\text{ToF}}}{c}, \quad (6)$$

where c is the speed of light. Estimation of the depth follows from the well-known 4-bucket trick, such that

$$\widehat{z_{\text{ToF}}} = \frac{c}{\Omega} \arctan \left(\frac{c(\frac{3\pi}{2}) - c(\frac{\pi}{2})}{c(0) - c(\pi)} \right). \quad (7)$$

Equations 2 through 7 summarize the general operation of many correlation ToF systems, sufficient for the context of this paper. However, for details on specific correlation ToF architectures, the reader is referred to review article by Masia [1] and Bhandari and Raskar [3].

B. THE FREQUENCY BANDWIDTH OF ToF CAMERAS

Increasing the frequency of correlation ToF camera frequencies from MHz to GHz offers benefits, but also poses some new challenges for machine vision applications.

1) HIGH-FREQUENCIES THEORETICALLY IMPROVE RANGE PRECISION

Increased range precision is perhaps the key motivation to use high frequencies for ToF imaging. It is clear from the form of Equation 3 that increasing the modulation frequency of phase ToF cameras offers greater depth precision. Using propagation of errors, Satat et al. [10] have derived the range accuracy ΔL as proportionate to

$$\Delta L \propto \frac{c}{\Omega}. \quad (8)$$

Let us define a quantity known as the “1 degree path”, representing the change in pathlength required to cause a phase shift of 1° in the measured signal. Expressed symbolically,

$$z_{1^\circ} = \frac{c\pi}{180\Omega}, \quad (9)$$

where z_{1° is the “1 degree path”. As illustrated in Table 1, the 1 degree path is much smaller at high-frequencies. A measurement scheme that can measure 1 degree phase shifts, would be able to obtain target range with a precision of 83.3 microns at a frequency of 10 GHz, but at 30 MHz, the range precision is only 27.8 millimeters. Figure 2 plots the pathlength change wrt. phase change at a variety of frequencies. In the context of this graph, z_{1° is the horizontal coordinate of the intersection of the dashed line with the sloped lines. To calculate depth precision, one would find the minimum, detectable phase shift of their instrumentation and find the curve intersection with the chosen modulation frequency.

TABLE 1. The z_{1° path for various ToF imagers at different frequencies.

Freq	Units	z_{1° path	Units	Device
30.0	MHz	27.8	mm	Mesa Swissr.
120	MHz	6.94	mm	MS Kinect
1.00	GHz	833	um	This paper
10.0	GHz	83.3	um	This paper

2) HIGH FREQUENCIES AID COMPUTATIONAL IMAGING APPLICATIONS

Correlation time of flight imagers have been used in computational imaging, realizing exotic applications like seeing around corners [21], or separating multipath interference [13], [16], [17], [38]–[40]. Several of these papers theoretically conclude that performance in these areas would improve if high-frequency ToF imaging was experimentally demonstrated. We highlight two recent papers that theoretically motivate the design of high-frequency imagers. First, Kadambi et al. [17] derive a global frequency bandlimit, which describes a minimum modulation frequency where global illumination vanishes. For tabletop scenes in computer

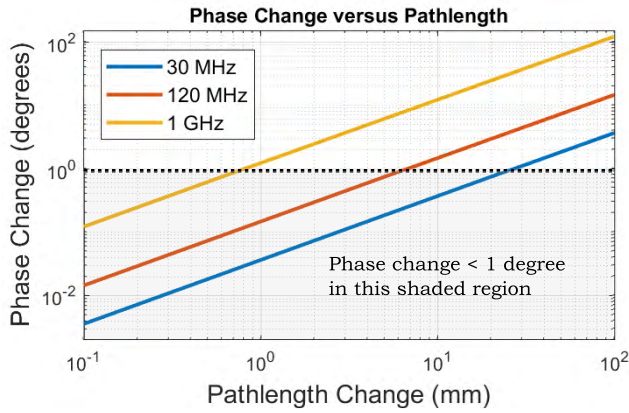


FIGURE 2. Measured phase depends on both frequency and pathlength. The shaded region of the plot indicates phase shifts less than 1 degree. The horizontal coordinate of the intersection of each colored line with the dashed line describes the pathlength change that causes a 1 degree phase change (hereafter, z_1° path).

vision and graphics, the global frequency bandlimit is on the order of 600 MHz to 6 GHz, which is not achievable with commodity ToF imagers today. More recently, Velten *et al.* [23] showed that the spatial resolution of recovering occluded scenes (e.g. around corners) increases with increasing camera modulation frequency. This paper proposes the design of a high-frequency ToF camera that could enhance such applications.

3) HIGH-FREQUENCIES CAUSE PHASE WRAPPING ARTIFACTS

As the frequency increases, the imaging system is susceptible to phase wrapping. Consider a concrete example where a phase ToF camera is operating at frequency Ω . Then, scene objects at a pathlength greater than

$$z_{\text{ambiguity}} = \frac{2c\pi}{\Omega} \quad (10)$$

will experience phase wrapping. However, it is possible to, separately, acquire data at a “nearby” frequency Ω^* , defined such that $|\Omega^* - \Omega| \ll \Omega$ and $|\Omega^* - \Omega| \ll \Omega^*$. Now, only scene

depths at a depth greater than

$$z_{\text{ambiguity}} = \frac{2c\pi}{\|\Omega - \Omega^*\|}, \quad (11)$$

will experience ambiguities at both Ω and Ω^* . It is true that this method of mitigating phase wrapping is an extra step, requiring measurements at an additional frequency. Nevertheless, the multiple frequency method is robust enough to find use in real-time consumer devices (e.g. the Microsoft Kinect). We must mention that the phase unwrapping methods may need to be modified to scale to GHz frequencies. The pioneering work of Gupta *et al.* [17] would serve as a starting point in this line of research.

4) WHEN IS THE FREQUENCY TOO HIGH?

Higher frequencies are not always desirable. Extremely high frequencies, on the order of THz or above, make the time of flight system hyper-sensitive to changes in pathlength. Environmental factors such as air turbulence or vibrations will induce a nuisance phase shift. Perhaps more importantly, at these frequencies the technique encroaches on the territory of optical interferometry (for which there are already existing solutions). We therefore consider THz frequency ranging out of scope for this paper.

5) WHAT LIMITS HIGH FREQUENCY ToF IMAGING?

The photosensor limits the highest frequency that existing correlation ToF imagers can operate at. It turns out that there is substantial headroom on the illumination side, as it is routine in optics and telecommunications applications to modulate laser light at frequencies upwards of 40 GHz. We exploit the frequency headroom in illumination, to send fast laser strobes to a target, which can be heterodyned down to a beat note within the photosensor bandwidth.

IV. SINGLE CASCADE

First we study a limiting case of cascaded ToF, where one modulator is placed in series with another demodulator. As shown in Figure 3, the field from a continuous-wave laser propagates through a Mach-Zehnder interferometer (denoted as $MZ\Omega_1$) with a crystal in one arm (hereafter,

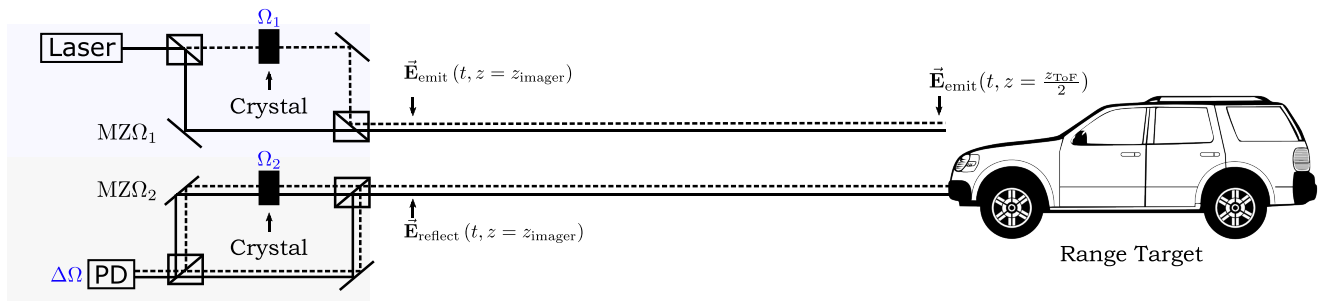


FIGURE 3. The optical schematic for a simplified image cascade. Light is emitted from a laser, and passes through Mach-Zehnder modulator $MZ\Omega_1$, which encodes a modulation signal on the laser light in the context of frequency Ω_1 . Light reaches a target at a range of $\frac{z_{\text{ToF}}}{2}$ away from the imager. The reflected field is passed through another Mach-Zehnder modulator $MZ\Omega_2$ with frequency context of Ω_2 . Since $\Omega_1 \neq \Omega_2$ a low-frequency beat note at $\Delta\Omega = \Omega_1 - \Omega_2$ is observed at the photodetector.

a “Mach-Zehnder modulator”). The potential difference across the crystal oscillates sinusoidally at a frequency of Ω_1 . This field is emitted at a range target. Before impinging upon the photodetector, the reflected light passes through another Mach-Zehnder modulator (denoted as $MZ\Omega_2$), where now Ω_2 denotes the voltage oscillation.

The forward model for this process can be shown to generate an (approximate) beat note, subject to approximation error (cf Appendix C). First, let us write a monochromatic plane wave as $E(t, z) = E_0 e^{j(\omega t - kz)}$, where z is the direction of propagation, E_0 is the amplitude of the field, ω is the angular frequency of optical radiation and $k = \frac{\omega}{c}$ is the wavenumber. To simplify notation, assume that z changes only as a result of path travel to the object. The electric field emitted toward the target has passed through the first Mach-Zehnder interferometer, such that

$$E_{\text{emit}}(t, z) = E_0 (e^{j(\omega t - kz)} e^{jn_0 + \eta \cos(\Omega_1 t - \Omega_1 \frac{z}{c})} + e^{j(\omega t - kz)}), \quad (12)$$

where $E_{\text{emit}}(t, z)$ denotes the field as emitted toward the target, n_0 denotes the index of refraction at zero voltage and η a constant of sensitivity of the crystal to voltage oscillations. In the context of time of flight range imaging, one is interested in three parameter values for z in Equation 12. These represent the

Field emitted from imager $E_{\text{emit}}(t, z = z_{\text{imager}})$

Field incident on target $E_{\text{emit}}(t, z = z_{\text{imager}} + \frac{z_{\text{ToF}}}{2})$

Field returning to imager $E_{\text{reflected}}(t, z = z_{\text{imager}})$, (13)

where, as shown on Figure 3, z_{imager} denotes the axial location of the emitted light and z_{ToF} represents the total pathlength traveled from the transmitter to receiver. The field that returns to the imager can be understood in context of the emitted field as

$$E_{\text{reflected}}(t, z = z_{\text{imager}}) = E_{\text{emit}}(t, z = z_{\text{ToF}}), \quad (14)$$

and expanded as

$$E_{\text{reflected}}(t, z = z_{\text{imager}}) = E_0 \underbrace{e^{j(\omega t - \omega \frac{z_{\text{ToF}}}{c})}}_{\text{optical term}} \times \underbrace{(e^{jn_0} e^{j\eta \cos(\Omega t - \Omega \frac{z_{\text{ToF}}}{c})})}_{\text{modulation term}} + 1), \quad (15)$$

The desired quantity from Equation 15 is the pathlength, z_{ToF} , which is encoded in the optical and modulation terms. The optical term is a sinusoid whose phase encodes the depth. Unfortunately, for macroscopic ranging, the phase term lacks meaning because the term is scaled by $\omega \approx 10^{15}$ Hz. Even air vibrations will cause an appreciable phase shift. Henceforth, we will not consider extracting z_{ToF} from the optical term. The goal is to recover z_{ToF} from the modulation term where z_{ToF} is scaled by $\Omega \approx 10^9$ Hz. Ω is an electronically controlled parameter that can be tuned using signal generation to sweep a range from DC to GHz.

The modulation term is non-sinusoidal (i.e. can be seen as a summation of Bessel functions) with GHz frequency harmonics, too fast for many photosensors. Therefore, this field is passed through a separate Mach-Zehnder to demodulate the signal, shown in Figure 3 as $MZ\Omega_2$. The demodulation Mach-Zehnder ($MZ\Omega_2$) follows the same structure as the modulation Mach-Zehnder ($MZ\Omega_1$), with one exception: the voltage signal is modulated at a frequency of $\Omega + \Delta\Omega$. The electric field at the output of $MZ\Omega_2$ then takes the form of

$$E_{\text{measurement}}^{\text{demod}} = E_0 e^{j(\omega t - \omega \frac{z_{\text{ToF}}}{c})} \left(e^{j \cos(\Omega t - \Omega \frac{z_{\text{ToF}}}{c})} + 1 \right) \times \left(1 + e^{j \cos((\Omega + \Delta\Omega)t)} \right) \quad (16)$$

Now, the field $E_{\text{measurement}}^{\text{demod}}$ is measured by the photodetector and approximated as:

$$I(t) \approx 0.88 E_0^2 \cos(-\Delta\Omega t - \underbrace{\Omega \frac{z_{\text{ToF}}}{c}}_{\varphi}). \quad (17)$$

The phase is estimated by correlating the low-frequency beat tone with a known reference. As is standard in correlation ToF imaging, let $r(t) = \cos(\Delta\Omega t)$ denote the reference signal. The correlation is therefore

$$c(\tau) = \lim_{T' \rightarrow \infty} \frac{1}{T'} \int_{-T'/2}^{T'/2} I(t) r(t + \tau) dt \approx 0.44 E_0^2 \cos(\Delta\Omega t + \Omega \frac{z_{\text{ToF}}}{c}) \quad (18)$$

The pathlength can now be estimated by using the relation

$$\widehat{z_{\text{ToF}}} = \frac{c}{\Omega} \arctan \left(\frac{c(\frac{3\pi}{2}) - c(\frac{\pi}{2})}{c(0) - c(\pi)} \right), \quad (19)$$

or any one of alternate approaches to extracting phase from correlation [3]. Intermediate derivations and error analysis of this process can be found in the appendix section of this paper.

V. MULTI CASCADE

The previous section described the limiting case where two modulators are used to obtain target range at GHz frequencies. However, there is a limit to the frequency of modulation to which an individual modulator can modulate the light. This is due to a limit on the frequency of voltage oscillations applied to the crystal. In practice this limit is based on the stability and quality of the crystal, but the highest performing LiNbO3 crystals support 100 GHz modulation on the upper end. Denote Ω_{crystal} as the maximum frequency with which the crystal can oscillate at.

We aim to generalize the imaging architecture from Section IV, to the case of any number of Mach-Zehnder modulators. To illuminate the scene with a frequency higher than Ω_{crystal} , we now exploit the high-frequency beat note.

Figure 4 illustrates the approach of stacked modulators. Denote $MZ\Omega_1^{(k)}$ as the k -th Mach-Zehnder prior to light illumination. The highest-frequency beat note is of primary

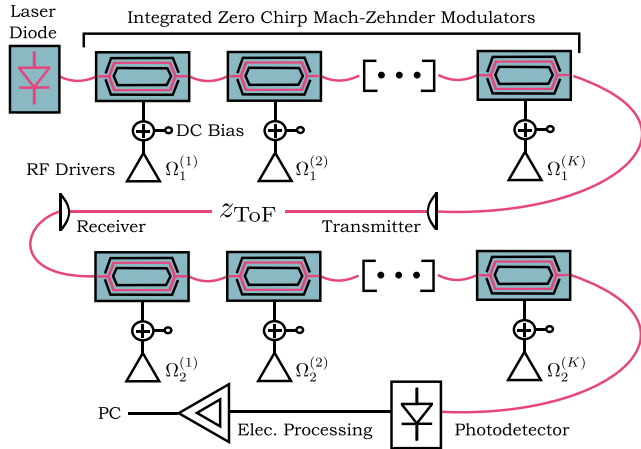


FIGURE 4. Generalizing Figure 3 to a multi-element cascade. Using $2K$ modulation elements enables frequency boosting and closer adherence to the sinusoidal approximation.

interest in the stacked configuration, which has a frequency of

$$\Omega_1^{\max} = \sum_{k=1}^K \Omega_1^{(k)}, \quad (20)$$

where Ω_1^{\max} represents the maximum modulation frequency of illumination emitted at the target. A similar approach of stacking Mach-Zehnder interferometers is used to demodulate the received signal, such that

$$\Omega_2^{\max} = \sum_{k=1}^K \Omega_2^{(k)}. \quad (21)$$

Since $\Omega_1^{\max} > \Omega_{\text{crystal}}$, it is possible to achieve even higher frequency modulation. When using such a cascade it is important to select the correct beat note. Only the highest-frequency beat note is shown in the summation equation in Figure 4; however, multiple combinations of beat notes are emitted at the target. Consider the case of $K = 3$ modulators used to illuminate the scene. The beat tones are

$$\Omega_1^{\text{stacked}} \rightarrow \left\{ \begin{array}{l} \Omega_1^{(1)} + \Omega_1^{(2)} + \Omega_1^{(3)} \\ \Omega_1^{(1)} + \Omega_1^{(2)} - \Omega_1^{(3)} \\ \Omega_1^{(1)} - \Omega_1^{(2)} + \Omega_1^{(3)} \\ \Omega_1^{(1)} - \Omega_1^{(2)} - \Omega_1^{(3)} \end{array} \right\}, \quad (22)$$

where the highest frequency, $\Omega_1^{(1)} + \Omega_1^{(2)} + \Omega_1^{(3)}$, would be of primary interest. To recover the highest frequency, the appropriate demodulation frequencies and narrowband detection are necessary to sense only the low-frequency beat note corresponding to the illumination frequency of interest.

VI. PERFORMANCE ANALYSIS

We compare the cascaded approach with existing alternatives in practical ranging scenarios.

A. ENVIRONMENTAL VIBRATIONS

Cascaded ToF is robust to vibrations because each pair of interferometric paths in the circuit are perturbed equally by a vibration (a form of common-mode rejection). In comparison, classical interferometry can be affected by vibrations that perturb interferometric paths differently.

In its original formulation in [6], superheterodyne interferometry was theorized to be more robust to vibrations than conventional interferometry. However, this is difficult to achieve in practice. For instance, a recent approach by Li *et al.* [9] demonstrated considerable experimental skill in transitioning superheterodyne interferometry to bulk fiberoptics, but as a bulk optical setup, is sensitive to environmental vibrations. Although this limitation is not fundamental; vibrations could be mitigated through engineering (e.g. by using polarization maintaining fiberoptics or through silicon implementation); cascaded modulation elements avoid this problem. A physical experiment demonstrating vibration tolerance can be found in the “Results” section of this paper.

B. BEAT FREQUENCY: DRIFT AND TUNING

Optical methods for heterodyne detection rely on using two lasers at a different optical frequency. The interference between these lasers exhibits a beat note at the difference in optical frequency. Unfortunately, it is difficult to maintain the frequency of the beat note. For example, suppose the optical frequency of a laser varies by one part in one million; at a wavelength of 1550nm, the frequency variation is around 2 MHz. Moreover, the frequency range of even tunable laser diodes optical radiation is limited.

In cascaded ToF, the beat note is generated through electronic modulation. The Mach-Zehnder modulator is a robust silicon photonic device that has been constructed for telecommunications applications. As such, the beat note does not drift in frequency using the cascaded approach. In fact, since the precision in the beat signal is less than 1 Hz, a stable beat frequency difference of less than 30 Hz can be imposed between the two modulators (motivating the application of using a video camera to obtain ToF measurements). A physical experiment demonstrating precision in the beat note can be found in the “Results” section of this paper.

C. NANOMETER RANGE PRECISION

Cascaded ToF addresses the GHz gap, but is not the preferred method for achieving THz frequencies, and correspondingly, nanometer range precision. While, in theory, cascaded ToF could stack many GHz modulators in sequence, both interferometry and superheterodyne interferometry offer a more elegant (and practical) solution toward achieving nanometer range precision.

In current experimental form, our physical experiments demonstrate range precision on the order of a micron (see the “Results” section).

D. LOW-COHERENCE AND LONG RANGING

Interferometry requires coherent light. The coherence length must be greater than the path difference between the arms. While this condition holds in microscopic interferometry, it is not the case for macroscopic interferometry, where objects may be many meters away. In superheterodyne interferometry, a long coherence laser source must be used. In comparison, cascaded ToF does not have strict requirements on the coherence as the object is present outside an interferometric path. Hence, the range of far away objects can be obtained as long as the laser coherence is longer than the microscopic path differences in the Mach-Zehnder modulators.

VII. RESULTS

To evaluate the cascaded method of high-resolution ranging, we conducted synthetic and physical experiments.

A. SYNTHETIC RESULTS

A computer graphics approach was used to mimic experimental hardware. 3D models are used as ground truth. To conduct accurate forward measurements we use the method of path tracing from computer graphics, allowing both direct and global illumination to be simulated accurately. We simulate read and shot noise in accordance with the simulations from prior work in the computer graphics and computer vision communities [16], [17].

1) SYNTHETIC DEPTH RECOVERY

Figure 5 illustrates a qualitative comparison of scanning a small, detailed object. The rendered, ground truth shape is a key approximately a centimeter in width and consisting of sub-mm features like the grooves. Reconstructed shapes using cascaded ToF at various frequencies are illustrated in Figure 5a-e. At a frequency of 10 GHz, the 3D surface of the key is reproduced faithfully. The reconstructions diminish in quality as the frequency is reduced. The plot in Figure 5 shows a horizontal slice through the recovered depth map. The plot shows that the grooves in the key are captured accurately by the 10 GHz waveform, but lower frequencies exhibit substantial noise. While it would be possible to smooth the low-frequency information from Kinect-like data, such an approach would also smooth out fine detail [41].

2) SYNTHETIC RESOLUTION CHART

To quantitatively address resolution we render a 3D resolution chart in Figure 6a. The chart consists of multiple raised elements at different heights from a background with features ranging in size from 50 microns to approximately 1 centimeter. As shown in Figures 6b and 6c, there is a qualitative difference in scans of the object at 500 MHz and at 10 GHz. A quantitative plot shows that the log of percent error in depth reconstruction varies as a function of feature size and frequency. As expected, smaller features and lower frequencies lead to greater errors.

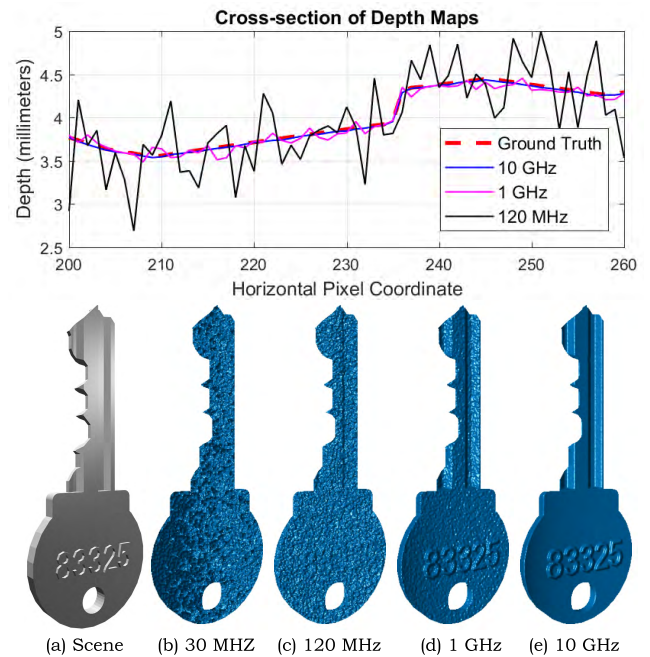


FIGURE 5. Comparing the cascaded GHz approach with Kinect-style approaches (simulated result). (a) The scene is a Yale key with sub-mm topographic features. (b) The simulated output from a 30 MHz ToF imager (e.g. Mesa Swissranger). (c) Simulated output from a 120 MHz ToF imager (e.g. Microsoft Kinect). (d) Simulated output from beat note ToF, where the modulation block is modulated at $\Omega_1 = 1$ GHz. (e) Simulated output from beat note ToF, where the modulation block is modulated at $\Omega_1 = 10$ GHz. (top row) A cross-section plot of the recovered depth maps for ground truth, 10 GHz (beat note ToF), 1 GHz (beat note ToF) and 120 MHz (Kinect simulation).

3) A MACHINE VISION COMPARISON WITH THE KINECT

Thus far, the scenes considered have been microscopic and highly detailed, where higher-frequencies have a compelling advantage. The aim of this paper is to transition telecommunication ideas to the realm of vision and graphics. In these communities, scans of human faces are of continued interest. Figure 7 shows a rendered scene of Albert Einstein, scaled to proper human proportions (3D model credit: Lloyd Chidgey). The model consists of both the large-scale contour of the head, but also includes fine-detail like facial wrinkles, hair and clothing folds. Since the scene is large, the cascaded approach at 500 MHz and 1 GHz are sufficient to obtain detailed scans (Figures 7a, 7b), while a Kinect-like camera exhibits noise in the regions of detail (Figure 7c).

B. EXPERIMENTAL RESULTS

To evaluate performance in machine vision conditions, a bulk fiberoptic setup is constructed (shown in Figure 8). The setup mimics the single cascade and uses a micrometer delay stage to precisely measure pathlength changes. Unless otherwise mentioned, the frequencies supplied to the two modulators are 3 GHz and 3 GHz + 10 Hz, creating a 10 Hz beat note.

1) CHARACTERIZING RANGE PRECISION

To measure range precision, 200 phase samples were extracted from the measurement device. These samples were

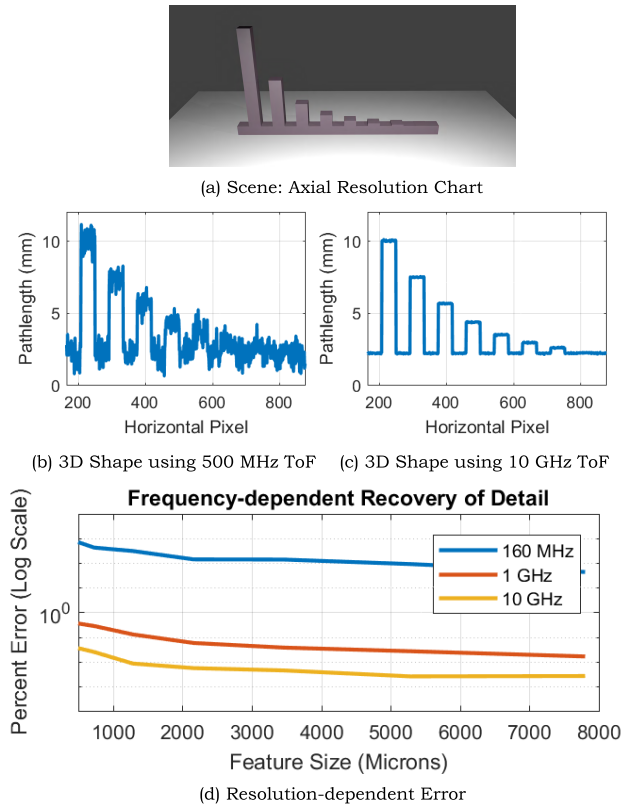


FIGURE 6. Path traced rendering of a resolution chart (simulated result) (a) A depth resolution chart consisting of raised surfaces at different heights from the background. (b) The recovered resolution chart at 1 GHz has some noise and smoothing while (c) the recovered chart at 10 GHz is a more faithful reproduction. The error for different levels of detail is plotted in (d), for three different modulation frequencies.

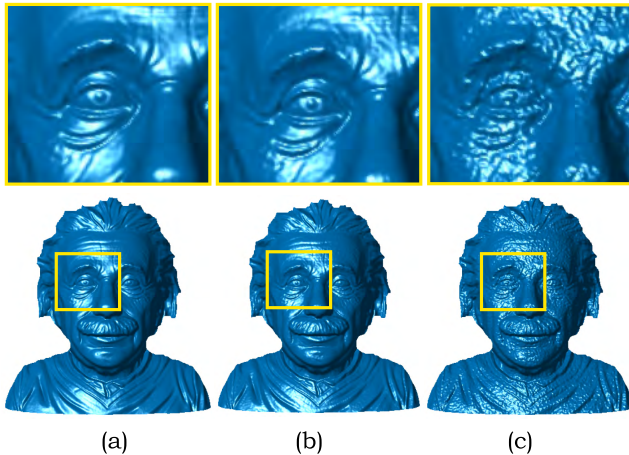


FIGURE 7. A GHz ToF imager achieves detailed human scans (simulated result). A bust of Albert Einstein scaled to human proportions is scanned using the cascaded approach at (a) 1 GHz and (b) 500 MHz. A Kinect simulating engine was used to scan the bust again at (c) 120 MHz.

then converted to a range measurement using the relation:

$$\widehat{z_{\text{ToF}}} = \frac{c}{\Omega} \varphi, \quad (23)$$

where φ is the phase reading from the electronics. As shown in Figure 9, over a thirty second measurement window, the range precision is under 3 microns. Over a longer acquisition

Testing Pathlength Precision with Micrometer Stage

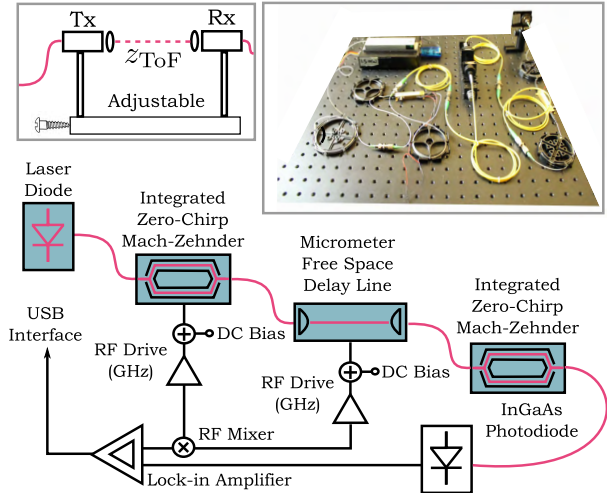


FIGURE 8. A micrometer resolution stage was used to evaluate the technique. The fiberoptic circuit is shown in the figure. Purple lines denote fiberoptic cable paths, and black lines denote electronic paths. The micrometer delay line is 5 centimeters long and has an adjustment screw to vary the distance between the transmit and receive fiber collimators. The upper-right inset shows the bulk optical setup.

of 10 minutes, the range precision is under 14 microns. Both figures are more precise than what correlation ToF imagers enable today.

2) CHARACTERIZING PERFORMANCE FOR MACHINE VISION

Machine vision applications are characterized by reflective loss, vibrations, and the use of low-framerate sensors. Performance for these three cases are illustrated in Figure 10. A variable optical attenuation filter was used to mimic losses in transmit power. As shown in the curve, the range precision is unaffected at up to 30 dB of power loss. The attenuation filter serves as a proxy for targets with low reflectivity or at long ranges.

The vibration test in Figure 10 was conducted by using a vibrating speaker to induce vibrations asymmetrically to the delay stage. A calibrated vibration probe was used to measure the vibrations. The induced vibrations were in the range of class A to class C vibrations, according to the ISO 2372 standard (class A vibrations are “minor” while class C vibrations are the “maximum tolerable”). The experimental results show that the range precision is not affected by the vibration velocity.

The beat note test in Figure 10 suggests that video framerate photodetectors could be used to conduct time of flight measurements. With the modulation scheme at 3 GHz, performance was tested over a range of beat notes from 1 Hz to 10 KHz. It is expected and confirmed through experimentation that higher frequencies (and shorter periods) show improved range precision. At 1 Hertz, the required frequency stability is about 1 part in 1 billion, so the range precision is around 100 microns. Interestingly, the knee of the curve is around 10 Hz, where the range precision is around

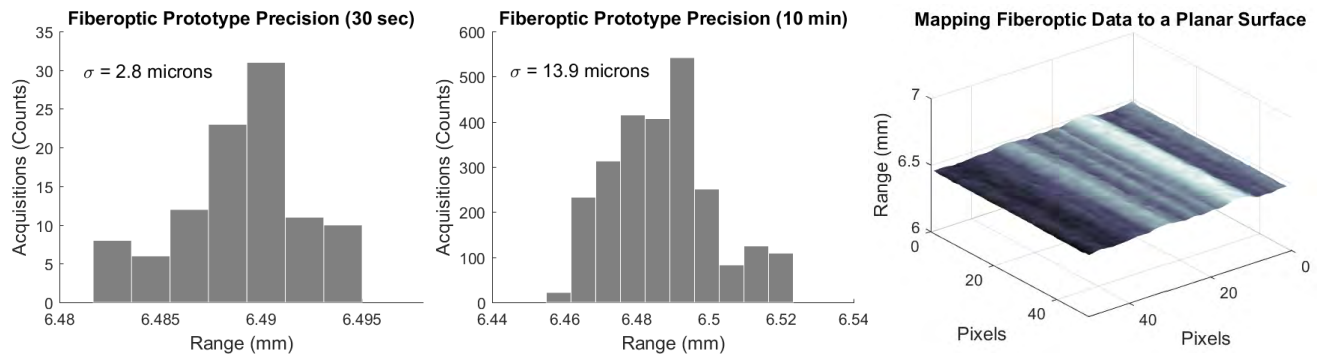


FIGURE 9. The prototype demonstrates a range precision of 2.8 microns (experimental result). (left) The standard deviation in range estimation at 3 GHz. (middle) Due to a non-random drift in the electronics, the range precision is 13.9 microns over a ten minute sampling window, which has implications when (right) scanning a planar surface.

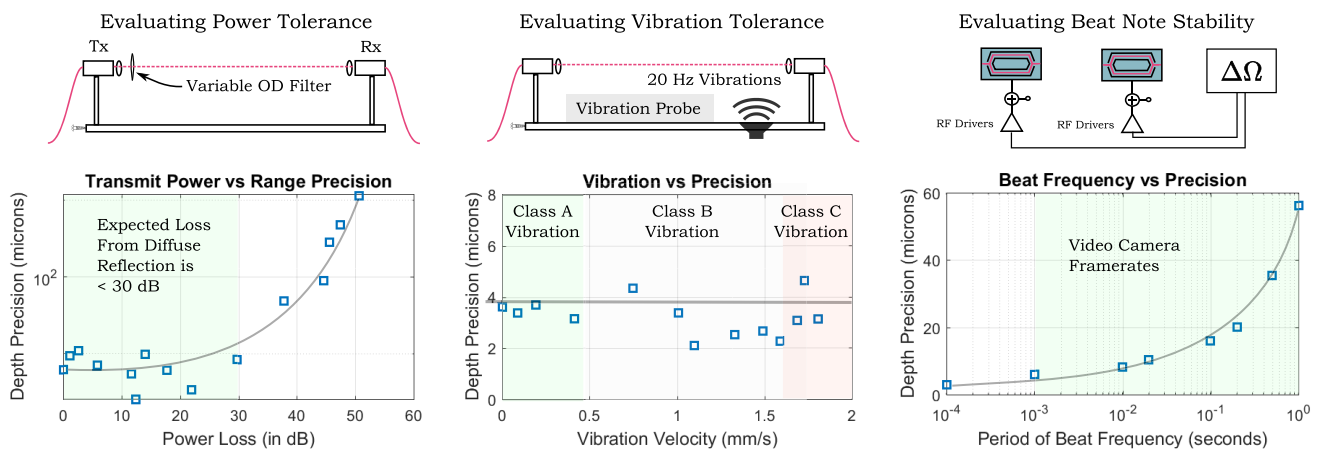


FIGURE 10. Characterizing range precision in challenging conditions. (left) Optical filters are placed in the path of the transmitted laser radiation. The range precision remains stable until the optical power is reduced by about 45 dB. The expected free space coupling loss is 30 dB. (middle) The range precision does not change with increasing vibrations. The vibration measurements are in context of the ISO 2372 standard. (right) The beat note is reduced to 1 Hertz for a base frequency of 3 GHz.

10 microns. Camera sensors which exceed the Nyquist frequency of 20 Hz are widely available.

3) LINEARITY IN PHASE AND FREQUENCY

For a given pathlength change, the measured phase shift is linearly scaled by the frequency. Figure 11 illustrates results

from an experiment where the micrometer delay stage is varied at 250 micron increments. Over a travel range of 3250 microns (or 3.25 mm), the change in phase at a base frequency of 30 MHz is negligible. In comparison, the change in phase at 7 GHz is nearly 30 degrees. The experimental phase shift observed between 3 GHz, 5 GHz, and 7 GHz frequencies are consistent with the expected result.

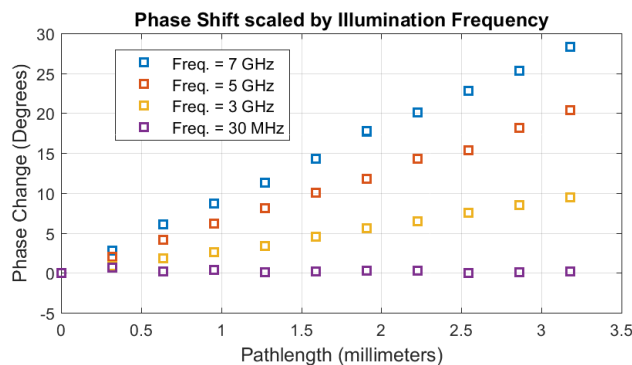


FIGURE 11. Evaluating linearity in the phase-frequency relationship (experimental result). This figure is an experimental evaluation of the synthetic example shown in Figure 2.

VIII. DISCUSSION

The experiments and simulations introduced in our paper suggest that a cascaded method could boost the frequency bandwidth of existing time of flight imagers. The cascaded method allows a single laser to be used with tunable frequency selection. However, there remains work to transition the fundamental ideas to a full-scale implementation that realizes the benefits of high-frequency time of flight imaging.

Our current experiments have reinforced the conspicuous link between GHz time of flight imaging and precise 3D sensing. However, we believe this is only one of the possible applications. Let us dive deeper into the potential application of imaging hidden objects around corners using correlation ToF imagers. Kadambi *et al.* [23] derived an analytical form

for the spatial resolution to which an image of a non-line-of-sight scene (e.g. around corners) could be recovered. The provided bound is written as

$$\text{Seeing around corner resolution} \approx \frac{2\pi c C_1}{2\pi + C_1 C_2 \Omega}, \quad (24)$$

where C_1 and C_2 represent constants for the reflectance and size of the visible scene. Hence, for the same scene, a 30 MHz ToF imager (as used by Kadambi *et al.* [23]) will spatially resolve the occluded scene to 1 meter (a very blurry image), while a 100 GHz imager would resolve the same scene to one centimeter resolution. There are many other computational imaging applications where high-frequency ToF imaging is important [16], [17], [42].

The experimental results in this paper are in context of a single-pixel ToF imagers. The single-pixel approach is useful and at times even preferable to full-field imagers (e.g. self-driving cars use point LIDAR in part to mitigate optical scattering). However, there are certainly many machine vision applications that demand wide-field imaging. We are currently exploring how the ideas described in this paper could be transitioned to wide-field. This is a non-trivial extension. Wide-field, modulating elements often have a low frequency bandwidth as compared to single pixel fiberoptic equipment. However, this is precisely where a cascaded approach might be most useful. We could imagine either an optical implementation of Figure 3 or a cascaded stack of microchannel plates as a way to obtain high-frequency range information from a videocamera. It is important to note that the cascaded approach should generalize to a number of modulating elements. These include micro-channel plates, polarization modulation, Pockel's cells, and LCD panels, to name a few.

IX. CONCLUSION

In conclusion, we have achieved micron-scale depth resolution by connecting correlation ToF imaging with a cascaded scheme of heterodyning. The primary experimental conclusions are: (1) depth sensing at 3 micrometer resolution; (2) at Hertz-level framerates; and (3) robustness to extreme vibrations. The current results appear poised to generalize to the field of computer vision and graphics, where high-quality 3D scans must be obtained in uncontrolled environments. We believe that the our study has laid a foundation for future work in computational imaging where cascaded modulating elements are incorporated into the correlation ToF architecture.

APPENDIX A

DERIVATION OF DEMODULATED FIELD

The simplification to obtain the electric field impinging on the detector is provided here. Write the field returning to the imager as

$$E_{\text{reflected}}(t, z = z_{\text{imager}}) = E_0 e^{j(\omega t - \omega \frac{2d}{c})} (e^{jn_0} e^{j\eta \cos(\Omega t - \Omega \frac{2d}{c})} + 1). \quad (25)$$

This field passes through the demodulating Mach-Zehnder interferometer, denoted as $MZ\Omega_2$ in Figure 3. The applied voltage field in the demodulation Mach-Zehnder is in the same form as the modulation Mach-Zehnder, taking the form of

$$V^{\text{demod}}(t, z) = \cos(\Omega^{\text{demod}} t - \Omega \frac{z}{c}). \quad (26)$$

In principle, the coordinate for z at the imager has been fixed to zero throughout this paper. Hence, the applied voltage field is parametric only in time, such that

$$V^{\text{demod}}(t) = \cos(\Omega^{\text{demod}} t). \quad (27)$$

The frequency of demodulation can be expressed as

$$\Omega^{\text{demod}} = \Omega + \Delta\Omega, \quad (28)$$

where $\Delta\Omega$ is, in practice, ranges from 10^0 to 10^3 Hz, within the measurement capabilities of available videocameras. Subsequent derivations for $MZ\Omega_2$ parallel those of $MZ\Omega_1$. First, the incident radiation is split into upper and lower fields, taking the form of

$$\begin{aligned} E_{\text{upper}}^{\text{demod}} &= E_0 e^{j(\omega t - \omega \frac{2d}{c})} (e^{jn_0} e^{j\eta \cos(\Omega t - \Omega \frac{2d}{c})} + 1) e^{jn_0} \\ &\quad \times e^{j(\eta \cos((\Omega + \Delta\Omega)t))} \\ E_{\text{lower}}^{\text{demod}} &= E_0 e^{j(\omega t - \omega \frac{2d}{c})} (e^{jn_0} e^{j\eta \cos(\Omega t - \Omega \frac{2d}{c})} + 1) \end{aligned} \quad (29)$$

The demodulated field then impinges on a photodetector. Refer to this field as the measurement field, written as

$$\begin{aligned} E_{\text{measurement}}^{\text{demod}} &= E_0 e^{j(\omega t - \omega \frac{2d}{c})} (e^{jn_0} e^{j\eta \cos(\Omega t - \Omega \frac{2d}{c})} + 1) \\ &\quad \times e^{jn_0} e^{j(\eta \cos((\Omega + \Delta\Omega)t))} + E_0 e^{j(\omega t - \omega \frac{2d}{c})} \\ &\quad \times (e^{jn_0} e^{j\eta \cos(\Omega t - \Omega \frac{2d}{c})} + 1) \\ &= E_0 e^{j(\omega t - \omega \frac{2d}{c})} (e^{jn_0} e^{j\eta \cos(\Omega t - \Omega \frac{2d}{c})} + 1) \\ &\quad \times (1 + e^{jn_0} e^{j(\eta \cos((\Omega + \Delta\Omega)t))}) \\ &= E_0 e^{j(\omega t - \omega \frac{2d}{c})} (e^{j\cos(\Omega t - \Omega \frac{2d}{c})} + 1) \\ &\quad \times (1 + e^{j(\cos((\Omega + \Delta\Omega)t))}), \end{aligned} \quad (30)$$

where the equality is a simplification from setting the constants $\eta = 1$ and $n_0 = 0$.

APPENDIX B

DERIVATION OF IMAGE FORMATION MODEL

The photodetector measures the time averaged intensity, expressed as

$$I = \frac{1}{T} \int_0^T E_{\text{measurement}}^{\text{demod}} \bar{E}_{\text{measurement}}^{\text{demod}} dt, \quad (31)$$

where $\bar{\cdot}$ indicates the complex conjugate. The expression from Equation 30 will be used to calculate intensity (hence $\eta = 1$ and $n_0 = 0$). To further compact notation, let us use the following simplifications:

$$\begin{aligned} A &\triangleq \cos(\Omega t - \Omega \frac{2d}{c}) \\ B &\triangleq \cos((\Omega + \Delta\Omega)t) \end{aligned} \quad (32)$$

Then the intensity can be derived as

$$\begin{aligned}
 I &= \frac{1}{T} \int_0^T E_0 e^{j(\omega t - \omega \frac{2d}{c})} (e^{jA} + 1)(e^{jB} + 1) \\
 &\quad \times E_0 e^{j(\omega t - \omega \frac{2d}{c})} (e^{-jA} + 1)(e^{-jB} + 1) dt \\
 &= \frac{1}{T} \int_0^T E_0^2 \cos^2(\omega t - \omega \frac{2d}{c}) (e^{jA} + 1)(e^{jB} + 1) \\
 &\quad \times (e^{-jA} + 1)(e^{-jB} + 1) dt \\
 &= \frac{1}{T} \int_0^T E_0^2 \cos^2(\omega t - \omega \frac{2d}{c}) (e^{jA} e^{-jA} + e^{jA} + e^{-jA} + 1) \\
 &\quad \times (e^{jB} e^{-jB} + e^{jB} + e^{-jB} + 1) dt \\
 &= \frac{1}{T} \int_0^T E_0^2 \cos^2(\omega t - \omega \frac{2d}{c}) (e^{jA} + e^{-jA} + 2) \\
 &\quad \times (e^{jB} + e^{-jB} + 2) dt \\
 &= \frac{1}{T} \int_0^T E_0^2 \cos^2(\omega t - \omega \frac{2d}{c}) (2 \cos(A) + 2)(2 \cos(B) + 2) dt \\
 &= \frac{4}{T} \int_0^T E_0^2 \cos^2(\omega t - \omega \frac{2d}{c}) (\cos(A) + 1)(\cos(B) + 1) dt \\
 &= \frac{4}{T} \int_0^T E_0^2 \cos^2(\omega t - \omega \frac{2d}{c}) \\
 &\quad \times (\cos(A) \cos(B) + \cos(A) + \cos(B) + 1) dt \quad (33)
 \end{aligned}$$

Suppose that the integration time T is on the order of milliseconds (e.g. like an ordinary camera). Then, the intensity takes the form of

$$I(t) = \frac{4}{T} E_0^2 \int_0^T \cos^2(\omega t - \omega \frac{2d}{c}) \cos(A) \cos(B) dt. \quad (34)$$

Backsubstitute for A and B to obtain

$$\begin{aligned}
 I(t) &= \frac{4}{T} E_0^2 \int_0^T \cos^2(\omega t - \omega \frac{2d}{c}) \\
 &\quad \times \underbrace{\cos(\cos(\Omega t - \Omega \frac{2d}{c}))}_{\text{Bessel functions due to MZ}\Omega_1} \underbrace{\cos(\cos((\Omega + \Delta\Omega)t))}_{\text{Bessel functions due to MZ}\Omega_2} dt, \quad (35)
 \end{aligned}$$

where the quantity within the underbrace represents a quantity that could be written as a sum of Bessel functions. Suppose a DC Bias corresponding to $\pi/2$ was added to the voltage signal for each modulator (also known as the quadrature point). Then the intensity takes the form of

$$\begin{aligned}
 I(t) &= \frac{4}{T} E_0^2 \int_0^T \cos^2(\omega t - \omega \frac{2d}{c}) \\
 &\quad \times \underbrace{\cos(\cos(\Omega t - \Omega \frac{2d}{c}) + \frac{\pi}{2})}_{\text{Bessel functions due to MZ1}} \underbrace{\cos(\cos((\Omega + \Delta\Omega)t))}_{\text{Bessel functions due to MZ2}} dt. \\
 &= \frac{4}{T} E_0^2 \int_0^T \cos^2(\omega t - \omega \frac{2d}{c}) \\
 &\quad \times \underbrace{\sin(\cos(\Omega t - \Omega \frac{2d}{c}))}_{\text{Bessel functions due to MZ1}} \underbrace{\sin(\cos((\Omega + \Delta\Omega)t))}_{\text{Bessel functions due to MZ2}} dt. \quad (36)
 \end{aligned}$$

Now, it is possible to invoke the Jacobi-Anger expansion for Bessel functions. Recall that the Jacobi-Anger formula is

$$\sin(\cos(\theta)) = -2 \sum_{n=1}^{\infty} (-1)^n J_{2n-1}(1) \cos((2n-1)\theta), \quad (37)$$

where J_n indicates a Bessel function of the n -th kind. To illustrate the main point, it will suffice to use the first-order approximation of the Jacobi-Anger formula in subsequent calculations, such that

$$\begin{aligned}
 \sin(\cos(\theta)) &\approx 2J_1(1) \cos(\theta), \\
 &\approx \cos(\theta), \quad (38)
 \end{aligned}$$

where the second approximation follows from $2J_1(1) = 0.88$. Using these approximations, the intensity can be written as

$$\begin{aligned}
 I(t) &= \frac{4}{T} E_0^2 \int_0^T \cos^2(\omega t - \omega \frac{2d}{c}) \\
 &\quad \times \cos(\Omega t - \Omega \frac{2d}{c}) \cos((\Omega + \Delta\Omega)t) dt \quad (39)
 \end{aligned}$$

By using the product-sum identity, it is possible to express the intensity as

$$\begin{aligned}
 I(t) &= \frac{2}{T} E_0^2 \int_0^T \cos^2(\omega t - \omega \frac{2d}{c}) \\
 &\quad \times \left[\cos((2\Omega + \Delta\Omega)t - \Omega \frac{2d}{c}) + \cos(-\Delta\Omega t - \Omega \frac{2d}{c}) \right] dt \\
 &= \frac{2}{T} E_0^2 \left[\int_0^T \cos^2(\omega t - \omega \frac{2d}{c}) \cos((2\Omega + \Delta\Omega)t - \Omega \frac{2d}{c}) dt \right. \\
 &\quad \left. + \int_0^T \cos^2(\omega t - \omega \frac{2d}{c}) \cos(-\Delta\Omega t - \Omega \frac{2d}{c}) dt \right]. \quad (40)
 \end{aligned}$$

Since, the integration time is much longer than the period of $\cos((2\Omega + \Delta\Omega)t)$, it follows that

$$I(t) = \frac{2}{T} E_0^2 \int_0^T \cos^2(\omega t - \omega \frac{2d}{c}) \cos(-\Delta\Omega t - \Omega \frac{2d}{c}) dt. \quad (41)$$

Let us now use the opposite logic to simplify further. The period of $\cos(\Delta\Omega t)$ can be on the order of a second, while T is on the order of a millisecond, permitting the approximation

$$\begin{aligned}
 I(t) &\approx \cos(-\Delta\Omega t - \Omega \frac{2d}{c}) \frac{2}{T} E_0^2 \int_0^T \cos^2(\omega t - \omega \frac{2d}{c}) dt. \\
 &= E_0^2 \cos(-\Delta\Omega t - \Omega \frac{2d}{c}) \quad (42)
 \end{aligned}$$

In summary, the image formation model takes the form of a low-frequency beat tone with a phase resolution of GHz. This allows commodity devices, such as video cameras, ToF cameras, or lock-in amplifiers to detect the beat tone and recover the phase.

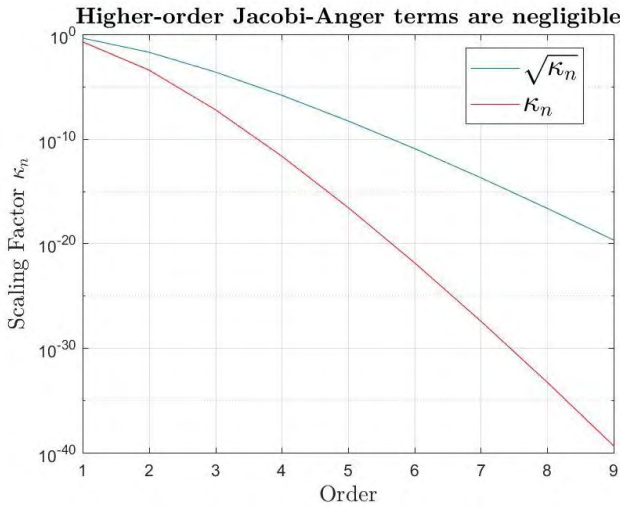


FIGURE 12. The approximation error from higher-order Jacob-Anger series terms decreases when using a cascaded approach. In the main paper, we neglected terms with order $n > 1$ in the Jacobi-Anger expansion. This is acceptable for cascaded ToF, as higher-order series terms decay rapidly in amplitude (blue curve), and even more rapidly for a cascade (red curve, 2-element cascade).

APPENDIX C

APPROXIMATION ERROR FROM SERIES EXPANSION

In deriving the image formation model, the first-order Jacobi-anger expansion was used to approximate the quantity $\sin(\cos(\theta))$ for some value of θ . Here, we show that the higher-order terms are negligible for a cascaded approach. The general form for the expansion of $\sin(\cos(\theta))$ is

$$\sin(\cos(\theta)) = -2 \sum_{n=1}^{\infty} (-1)^n J_{2n-1}(1) \cos((2n-1)\theta). \quad (43)$$

The image formation model prior to the expansion is

$$I(t) = \frac{4}{T} E_0^2 \int_0^T \cos^2(\omega t - \omega \frac{2d}{c}) \times \underbrace{\sin(\cos(\Omega t - \Omega \frac{2d}{c}))}_{\text{Bessel functions due to MZ1}} \underbrace{\sin(\cos((\Omega + \Delta\Omega)t))}_{\text{Bessel functions due to MZ2}} dt. \quad (44)$$

Using the Jacobi-Anger expansion yields:

$$I(t) = \frac{4}{T} E_0^2 \left[\int_0^T \cos^2(\omega t - \omega \frac{2d}{c}) \times \cos(\Omega t - \Omega \frac{2d}{c}) \times \cos((\Omega + \Delta\Omega)t) dt + \int_0^T \cos^2(\omega t - \omega \frac{2d}{c}) \times -2 \sum_{n=1}^{\infty} (-1)^n J_{2n-1}(1) \cos((2n-1)\Omega t - \Omega \frac{2d}{c}) \times -2 \sum_{n=1}^{\infty} (-1)^n J_{2n-1}(1) \cos((2n-1)(\Omega + \Delta\Omega)t) dt \right]. \quad (45)$$

Here, the value of the n -th approximation is scaled by J_{2n-1} . In particular, since the cascaded approach multiplies no fewer than two Jacobi-Anger expansions, the integral contributing to order n approximations is scaled by the square of J_{2n-1} (and a higher power for cascades of greater order K). More precisely, let κ denote the scaling constant of the expansions, where

$$\kappa_n = J_{2n-1}^2 \quad (46)$$

The scaling factor for the first order expansion $n = 1$ is about 0.2, but drops drastically for $n = 2$, to a value of 4×10^{-4} . Please refer to Figure 12 for details.

Acknowledgments

The authors thank Ayush Bhandari, Jack Erdozain, Barmak Heshmat, Anat Levin, Tomohiro Maeda, Yoav Schechner and Andy Tso for useful discussions. Further, the authors are grateful to the laboratories of Michael Watts (MIT RLE) and Mounji Bawendi (MIT Dept. of Chemistry) for loaning equipment that enabled this study.

REFERENCES

- [1] B. Masia, "Computational photography: Coding in space and time," in *Proc. 30th Spring Conf. Comput. Graph.*, 2014, pp. 115–121.
- [2] G. Satat, B. Heshmat, N. Naik, A. Redo-Sanchez, and R. Raskar, "Advances in ultrafast optics and imaging applications," *Proc. SPIE*, vol. 9835, p. 98350Q, May 2016. [Online]. Available: <https://www.spiedigitallibrary.org/conference-proceedings-of-spie/9835/1/Advances-in-ultrafast-optics-and-imaging-applications/10.1117/12.222438.full?SSO=1>
- [3] A. Bhandari and R. Raskar, "Signal processing for time-of-flight imaging sensors: An introduction to inverse problems in computational 3-D imaging," *IEEE Signal Process. Mag.*, vol. 33, no. 5, pp. 45–58, Sep. 2016.
- [4] A. Kolb, E. Barth, R. Koch, and R. Larsen, "Time-of-flight cameras in computer graphics," *Comput. Graph. Forum*, vol. 29, no. 1, pp. 141–159, 2010.
- [5] A. Payne *et al.*, "7.6 A 512 × 424 CMOS 3D time-of-flight image sensor with multi-frequency photo-demodulation up to 130 MHz and 2 GS/s ADC," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2014, pp. 134–135.
- [6] R. Dändliker, R. Thalmann, and D. Prongué, "Two-wavelength laser interferometry using superheterodyne detection," *Opt. Lett.*, vol. 13, no. 5, pp. 339–341, 1988.
- [7] R. Dändliker, M. Geiser, C. Giunti, S. Zatti, and G. Margheri, "Improvement of speckle statistics in double-wavelength superheterodyne interferometry," *Appl. Opt.*, vol. 34, no. 31, pp. 7197–7201, 1995.
- [8] Y. Tanaka, S. Tominaka, and T. Kurokawa, "Precision distance measurement using a two-photon absorption process in a silicon avalanche photodiode with saw-tooth phase modulation," *Appl. Opt.*, vol. 54, no. 28, pp. E35–E40, 2015.
- [9] F. Li, J. Yablon, A. Velten, M. Gupta, and O. Cossairt, "High-depth-resolution range imaging with multiple-wavelength superheterodyne interferometry using 1550-nm lasers," *Appl. Opt.*, vol. 56, no. 31, pp. H51–H56, 2017.
- [10] R. Lange and P. Seitz, "Solid-state time-of-flight range camera," *IEEE J. Quantum Electron.*, vol. 37, no. 3, pp. 390–397, Mar. 2001.
- [11] J. P. Godbaz, M. J. Cree, and A. A. Dorrington, "Mixed pixel return separation for a full-field ranger," in *Proc. 23rd Int. Conf. Image Vis. Comput. New Zealand (IVCNZ)*, Nov. 2008, pp. 1–6.
- [12] A. Dorrington, J. P. Godbaz, M. J. Cree, A. D. Payne, and L. V. Streeter, "Separating true range measurements from multipath and scattering interference in commercial range cameras," *Proc. SPIE*, vol. 7864, p. 786404, Dec. 2011. [Online]. Available: <https://www.spiedigitallibrary.org/conference-proceedings-of-spie/7864/1/Separating-true-range-measurements-from-multi-path-and-scattering-interference/10.1117/12.876586.full>

- [13] A. Bhandari, M. Feigin, S. Izadi, C. Rhemann, M. Schmidt, and R. Raskar, "Resolving multipath interference in Kinect: An inverse problem approach," in *Proc. IEEE SENSORS*, Nov. 2014, pp. 614–617.
- [14] D. Freedman, Y. Smolin, E. Krupka, I. Leichter, and M. Schmidt, "SRA: Fast removal of general multipath for ToF sensors," in *Proc. ECCV*, 2014, pp. 234–249.
- [15] 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 *Proc. ICASSP*, May 2014, pp. 365–369.
- [16] A. Kadambi, J. Schiel, and R. Raskar, "Macroscopic interferometry: Rethinking depth estimation with frequency-domain time-of-flight," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 893–902.
- [17] M. Gupta, S. K. Nayar, M. B. Hullin, and J. Martin, "Phasor imaging: A generalization of correlation-based time-of-flight imaging," *ACM Trans. Graph.*, vol. 34, no. 5, 2014, Art. no. 156.
- [18] F. Heide, M. B. Hullin, J. Gregson, and W. Heidrich, "Low-budget transient imaging using photonic mixer devices," *ACM Trans. Graph.*, vol. 32, no. 4, 2013, Art. no. 45.
- [19] J. Lin, Y. Liu, M. B. Hullin, and Q. Dai, "Fourier analysis on transient imaging with a multifrequency time-of-flight camera," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 3230–3237.
- [20] A. Velten *et al.*, "Femto-photography: Capturing and visualizing the propagation of light," *ACM Trans. Graph.*, vol. 32, no. 4, 2013, Art. no. 44.
- [21] 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 Commun.*, vol. 3, Mar. 2012, Art. no. 745.
- [22] F. Heide, L. Xiao, W. Heidrich, and M. B. Hullin, "Diffuse mirrors: 3D reconstruction from diffuse indirect illumination using inexpensive time-of-flight sensors," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 3222–3229.
- [23] A. Kadambi, H. Zhao, B. Shi, and R. Raskar, "Occluded imaging with time-of-flight sensors," *ACM Trans. Graph.*, vol. 35, no. 2, 2016, Art. no. 15.
- [24] R. M. Conroy, A. A. Dorrington, R. Kunemeyer, and M. J. Cree, "Range imager performance comparison in homodyne and heterodyne operating modes," *Proc. SPIE*, vol. 7239, p. 723905, Dec. 2009. [Online]. Available: <https://spie.org/Publications/Proceedings/Paper/10.1117/12.806139>
- [25] A. Kadambi, A. Bhandari, R. Whyte, A. Dorrington, and R. Raskar, "Demultiplexing illumination via low cost sensing and nanosecond coding," in *Proc. IEEE Int. Conf. Comput. Photogr.*, May 2014, pp. 1–9.
- [26] F. Heide, W. Heidrich, M. Hullin, and G. Wetzstein, "Doppler time-of-flight imaging," *ACM Trans. Graph.*, vol. 34, no. 4, 2015, Art. no. 36.
- [27] S. Shrestha, F. Heide, W. Heidrich, and G. Wetzstein, "Computational imaging with multi-camera time-of-flight systems," *ACM Trans. Graph.*, vol. 35, no. 4, 2016, Art. no. 33.
- [28] M. Bashkansky, H. R. Burris, E. E. Funk, R. Mahon, and C. I. Moore, "RF phase-coded random-modulation LIDAR," *Opt. Commun.*, vol. 231, no. 1, pp. 93–98, 2004.
- [29] S. Gao and R. Hui, "Frequency-modulated continuous-wave lidar using I/Q modulator for simplified heterodyne detection," *Opt. Lett.*, vol. 37, no. 11, pp. 2022–2024, 2012.
- [30] S. A. Benton and V. M. Bove, Jr., *Holographic Imaging*. Hoboken, NJ, USA: Wiley, 2008.
- [31] I. Gkioulekas, A. Levin, F. Durand, and T. Zickler, "Micron-scale light transport decomposition using interferometry," *ACM Trans. Graph.*, vol. 34, no. 4, 2015, Art. no. 37.
- [32] G. Satat, C. Barsi, B. Heshmat, D. Raviv, and R. Raskar, "Locating fluorescence lifetimes behind turbid layers non-invasively using sparse, time-resolved inversion," in *Proc. CLEO, Sci. Innov. Opt. Soc. Amer.*, 2014, pp. 1–2, paper JTh2A-43.
- [33] G. Satat *et al.*, "Locating and classifying fluorescent tags behind turbid layers using time-resolved inversion," *Nature Commun.*, vol. 6, Apr. 2015, Art. no. 6796.
- [34] G. Gariépy *et al.*, "Single-photon sensitive light-in-flight imaging," *Nature Commun.*, vol. 6, Jan. 2015, Art. no. 6021.
- [35] G. Gariépy, F. Tonolini, R. Henderson, J. Leach, and D. Faccio, "Detection and tracking of moving objects hidden from view," *Nature Photon.*, vol. 10, no. 1, pp. 23–26, 2016.
- [36] M. O'Toole, F. Heide, D. Lindell, K. Zang, S. Diamond, and G. Wetzstein, "Reconstructing transient images from single-photon sensors," in *Proc. IEEE CVPR*, Jul. 2017, pp. 2289–2297.
- [37] G. Satat, M. Tancik, O. Gupta, B. Heshmat, and R. Raskar, "Object classification through scattering media with deep learning on time resolved measurement," *Opt. Exp.*, vol. 25, no. 15, pp. 17466–17479, 2017.
- [38] A. Kadambi *et al.*, "Coded time of flight cameras: Sparse deconvolution to address multipath interference and recover time profiles," *ACM Trans. Graph.*, vol. 32, no. 6, 2013, Art. no. 167.
- [39] 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 Trans. Graph.*, vol. 33, no. 4, 2014, Art. no. 87.
- [40] 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," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 73–81.
- [41] A. Kadambi, V. Taamazyan, B. Shi, and R. Raskar, "Polarized 3D: High-quality depth sensing with polarization cues," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 3370–3378.
- [42] A. Bhandari, C. Barsi, and R. Raskar, "Blind and reference-free fluorescence lifetime estimation via consumer time-of-flight sensors," *Optica*, vol. 2, no. 11, pp. 965–973, 2015.



ACHUTA KADAMBI is currently working toward the Ph.D. degree at the Massachusetts Institute of Technology, Cambridge, MA, USA. His research is in computational imaging, a field at the intersection of computer science, and optics. His publications have been presented at CVPR, SIGGRAPH, ICCV, and ICCP. He is active on the program committees of various conferences and has co-organized a tutorial on time of flight imaging at ICCV and courses on 3-D imaging at SIGGRAPH.

He is listed as a co-inventor on over 15 U.S. patent applications. He is currently co-authoring a textbook on time of flight imaging (MIT Press, expected 2018).



RAMESH RASKAR is currently an Associate Professor at the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA. He joined the MIT Media Laboratory from the Mitsubishi Electric Research Laboratories in 2008 as the Head of the Laboratories Camera Culture Research Group. His research interests span the fields of computational photography, inverse problems in imaging, and human-computer interaction. Recent inventions include transient imaging to look around a corner,

a next-generation CAT-scan machine, imperceptible markers for motion capture (Prakash), long-distance barcodes (Bokode), touch and hover 3-D interaction displays (BiDi screen), low-cost eye care devices (NETRA), and new theoretical models to augment light fields to represent wave phenomena. In 2003, he received the Global Indus Technovator Award at MIT to recognize the top 20 Indian technology innovators worldwide, and the TR100 Award from Technology Review, presented to top young innovators under the age of 35, in 2004. In 2009, he received a Sloan Research Fellowship. In 2010, he received the DARPA Young Faculty Award. He holds over 75 U.S. patents, and has received four Mitsubishi Electric Invention Awards. He is currently coauthoring a book on computational photography.

...