CS-ToF: High-resolution compressive time-of-flight imaging

FENGQIANG LI,[1,3,*] HUAIJIN CHEN,[2,3] ADITHYA PEDIREDLA,[2] CHIAKAI YEH,[1] KUAN HE,[1] ASHOK VEERARAGHAVAN,[2] AND OLIVER COSSAIRT[1]

[1] *Department of Electrical Engineering and Computer Science, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208, USA*
[2] *Department of Eletrical and Computer Engineering, Rice University, 6100 Main Street, Houston, TX 77005, USA*
[3] *These authors contributed equally to this work*
**fengqiangli2015@u.northwestern.edu*

Abstract: Three-dimensional imaging using Time-of-flight (ToF) sensors is rapidly gaining widespread adoption in many applications due to their cost effectiveness, simplicity, and compact size. However, the current generation of ToF cameras suffers from low spatial resolution due to physical fabrication limitations. In this paper, we propose CS-ToF, an imaging architecture to achieve high spatial resolution ToF imaging via optical multiplexing and compressive sensing. Our approach is based on the observation that, while depth is non-linearly related to ToF pixel measurements, a phasor representation of captured images results in a linear image formation model. We utilize this property to develop a CS-based technique that is used to recover high resolution 3D images. Based on the proposed architecture, we developed a prototype 1-megapixel compressive ToF camera that achieves as much as 4 × improvement in spatial resolution and 3 × improvement for natural scenes. We believe that our proposed CS-ToF architecture provides a simple and low-cost solution to improve the spatial resolution of ToF and related sensors.

OCIS codes: (150.0150) Machine vision; (110.6880) Three-dimensional image acquisition; (100.6640) Superresolution; (100.6890) Three-dimensional image processing; (100.3190) Inverse problems.

References and links

1. J. Levinson, J. Askeland and J. Becker, J. Dolson, D. Held, S. Kammel, J. Z. Kolter, D. Langer, O. Pink, V. Pratt, M. Sokolsky, G. Stanek, D. Stavens, A. Teichman, M. Werling, and S. Thrun, "Towards fully autonomous driving: Systems and algorithms," in *IEEE Intelligent Vehicles Symposium* (2011), pp. 163–168.
2. K. Gallo and G. Assanto, "Vision based obstacle detection for wheeled robots," in *International Conference on Control, Automation and Systems* (2008), pp. 1587–1592.
3. G. A. Howland, D. J. Lum, M. R. Ware, and J. C. Howell, "Photon counting compressive depth mapping," Opt. Express **21**(20), 23822–23837 (2013).
4. L. R. Bissonnette, *Lidar: Range-Resolved Optical Remote Sensing of the Atmosphere* (Springer, 2005).
5. C. D. Mutto, P. Zanuttigh, and G. M. Cortelazzo, "Time-of-flight cameras and microsoft kinect," (Springer, 2012).
6. S. Foix, G. Alenya, and C. Torras, "Lock-in time-of-flight (ToF) cameras: a survey," IEEE Sens. J. **11**(9), 1917–1926 (2011).
7. R. Lange and P. Seitz, "Solid-state time-of-flight range camera," IEEE J. Quantum Electron. **37**(3), 390–397 (2001).
8. Microsoft Kinect Sensor (The second verison), https://developer.microsoft.com/en-us/windows/kinect, (2016).
9. TI ToF sensor, https://www.ti.com/sensing-products/optical-sensors/3d-time-of-flight/overview.html, (2016).
10. R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision* (Cambridge University, 2003).
11. C. Yeh, N. Matsuda, X. Huang, F. Li, M. Walton, and O. Cossairt, "A Streamlined Photometric Stereo Framework for Cultural Heritage," in *Proc. ECCV* (Springer, 2016), pp. 738–752.
12. F. Heide, M. Hullin, J. Gregson, and W. Heidrich, "Low-budget transient imaging using photonic mixer devices," ACM Trans. Graph. **32**(4), 45 (2013).
13. 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 Trans. Graph. **32**(6), 167 (2013).
14. F. Heide, W. Heidrich, M. Hullin, and G. Wetzstein, "Doppler time-of-flight imaging," ACM Trans. Graph. **34**(4), 36 (2015).

#303963 https://doi.org/10.1364/OE.25.031096
Journal © 2017 Received 21 Aug 2017; revised 16 Oct 2017; accepted 17 Nov 2017; published 29 Nov 2017

15. F. Heide, L. Xiao, A. Kolb, M. Hullin, and W. Heidrich, "Imaging in scattering media using correlation image sensors and sparse convolutional coding," Opt. Express **22**(2), 26338–26350 (2014).
16. K. Yasutomi, T. Usui, S. Han, M. Kodama, T. Takasawa, K. Kagawa, and S. Kawahito, "A time-of-flight image sensor with sub-mm resolution using draining only modulation pixels," in *Proc. Int. Image Sensor Workshop* (2013), pp. 357–360.
17. S. Kim, S. Cha, H. Park, J. Gong, Y. Noh, W. Kim, S. Lee, D.-K. Min, W. Kim, and T.-C. Kim, "Time of flight image sensor with 7um pixel and 640× 480 resolution," in *IEEE Symposium on VLSI Technology* (2013), T146–T147.
18. S. Schuon, C. Theobalt, J. Davis, and S. Thru, "High-quality scanning using time-of-flight depth superresolution," in *Proc. CVPR Workshops* (IEEE, 2008), pp. 1–7.
19. R. Nair, K. Ruhl, F. Lenzen, S. Meister, H. Schafer, C. S. Garbe, M. Eisemann, M. Magnor, and D. Kondermann, "A survey on time-of-flight stereo fusion," Springer, 105–127 (2013).
20. G. D. Evangelidis, M. Hansard, and R. Horaud, "Fusion of range and stereo data for high-resolution scene-modeling," IEEE Trans. Pattern Anal. Mach. Intell. **37**(11), 2178–2192 (2015).
21. C. Ti, R. Yang, J. Davis, and Z. Pan, "Simultaneous Time-of-Flight Sensing and Photometric Stereo With a Single ToF Sensor," in *Proc. CVPR* (IEEE, 2015), pp. 4334–4342.
22. A. Kadambi, V. Taamazyan, B. Shi, and R. Raskar, "Polarized 3d: High-quality depth sensing with polarization cues," in *Proc. CVPR* (IEEE, 2015), pp. 3370–3378.
23. L. Xiao, F. Heide, M. O'Toole, A. Kolb, M. B. Hullin, K. Kutulakos, and W. Heidrich, "Defocus deblurring and superresolution for time-of-flight depth cameras," in *Proc. CVPR* (IEEE, 2015), pp. 2376–2384.
24. J. Xie, C.-C. Chou, R. Feris, and M.-T. Sun, "Single depth image super resolution and denoising via coupled dictionary learning with local constraints and shock filtering," in *Proc. ICME* (IEEE, 2014), pp. 1–6.
25. J. Xie, R. S. Feris, and M.-T. Sun, "Edge-guided single depth image super resolution," IEEE Trans. Image Process. **25**(1), 428–438 (2016).
26. X. Song, Y. Dai, and X. Qin, "Deep Depth Super-Resolution: Learning Depth Super-Resolution using Deep Convolutional Neural Network," arXiv, (2016).
27. D. W. Davies, "Spatially multiplexed infrared camera," J. Opt. Soc. Am. **65**(6), 707–711 (1975).
28. R. N. Ibbett, D. Aspinall, and J. F. Grainger, "Real-time multiplexing of dispersed spectra in any wavelength region," Appl. Opt. **7**(6), 1089–1094 (1968).
29. N. J. A. Sloane, T. Fine, P. G. Phillips, and M. Harwit, "Codes for multiplex spectrometry," Appl. Opt. **8**(10), 2103–2106 (1969).
30. J. A. Decker and M. Harwit, "Experimental operation of a Hadamard spectrometer," Appl. Opt. **8**, 2552 (1969).
31. E. D. Nelson and M. L. Fredman, "Hadamard spectroscopy," J. Opt. Soc. Am. **60**(12), 1664–1669 (1970).
32. Y. Y. Schechner, S. K. Nayar, and P. N. Belhumeu, "Multiplexing for optimal lighting," IEEE Trans. Pattern Anal. Mach. Intell. **29**(8), 1339–1354 (2007).
33. M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single-pixel imaging via compressive sampling," IEEE Signal Process. Mag. **25**(2), 83–91 (2008).
34. G. A. Howland, P. B. Dixon, and J. C. Howell, "Photon-counting compressive sensing laser radar for 3D imaging," Appl. Opt. **50**(31), 5917–5920 (2011).
35. A. Kirmani, A. Colaço, F. N. C. Wong, and V. K. Goyal, "Exploiting sparsity in time-of-flight range acquisition using a single time-resolved sensor," Opt. Express **19**(22), 21485–21507 (2011).
36. A. C. Sankaranarayanan, P. Turaga, R. Baraniuk, and R. Chellappa, "Compressive acquisition of dynamic scenes," in *Proc. ECCV* (Springer, 2010).
37. A. C. Sankaranarayanan, C. Studer, and R. G. Baraniu, "CS-MUVI: Video compressive sensing for spatial-multiplexing cameras," in *Proc. ICCP* (IEEE, 2012).
38. J. Wang, M. Gupta, and A. C. Sankaranarayanan, "LiSens: A scalable architecture for video compressive sensing," in *Proc. ICCP* (IEEE, 2015).
39. H. Chen, M. Salman Asif, A. C. Sankaranarayanan, and A. Veeraraghavan, "FPA-CS: Focal plane array-based compressive imaging in short-wave infrared," in *Proc. CVPR* (IEEE, 2015), pp. 2358–2366.
40. 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. **34**(5), 156 (2015).
41. 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. **33**(4), 87 (2014).
42. J. M. Bioucas-Dias and M. A. Figueiredo, "A new TwIST: Two-step iterative shrinkage/thresholding algorithms for image restoration," IEEE Trans. Image Process. **16**(12), 2992–3004 (2007).
43. Y. Endo, T. Shimobaba, T. Kakue, and T. Ito, "GPU-accelerated compressive holography" Opt. Express **24**(8), 8437-8445 (2016).
44. S. Yu, A. Khwaja, and J. Ma, "Compressed sensing of complex-valued data," Signal Processing **92**(2), 357–362 (2012).
45. D. Mittleman, "Sensing with terahertz radiation," Springer **85**, (2013).
46. J. H. Ender, "On compressive sensing applied to radar," Signal Processing **90**(5), 1402–1414 (2010).
47. D. Scharstein, and R. Szeliski, "High-accuracy stereo depth maps using structured light," in *Proc. CVPR* (IEEE, 2003), pp. I–I.

1. Introduction

Three-dimensional (3D) sensors are being adopted in a number of commercial applications including self-driving cars [1] and robotics [2]. In this paper, we focus on 3D Time-of-Flight (ToF) based sensors. ToF cameras can be broadly classified into two categories based on the illumination signal: pulsed/LIDAR [3,4] and continuous wave amplitude modulated (CWAM) ToF cameras, also referred as lock-in ToF cameras [5,6]. In this paper, we focus on CWAM-ToF and use the abbreviation 'ToF' to describe them for the remainder of this paper. ToF cameras [7–9] are a practical and promising approach for inexpensive active 3D sensing with range independent depth resolution (as compared to stereo or multiview triangulation [10, 11]) and compact form factor (as compared to light detection and ranging (LIDAR) devices [4]). In the last decade, other novel imaging applications using ToF cameras have also been developed. Heide et al. [12] and Kadambi et al. [13] created transient captures of light with ToF camera systems based on photonic mixer devices. Heide et al. [14] designed a doppler time-of-flight system that can compute the 3D velocity of the objects instead of depth. A fundamental limit of performance in all these applications is the low spatial resolution that is achieved [7, 15].

Why are ToF sensors low resolution? A ToF imager is a focal plane array that simultaneously encodes the scene's intensity and depth information at each pixel. ToF cameras typically consist of an amplitude modulated light source that actively illuminates the scene and is coupled with a correlation sensor at each pixel that is locked-in to the same frequency. Multiple measurements are obtained with different amount of phase shift between transmitted and detected light. The amplitude modulation in most ToF cameras is performed at a modulation frequency in the 10-100 MHz range and this frequency controls both the unambiguous range of depths and the depth resolution of the ToF sensor. Additional electronics is required to implement the correlation measurement individually at each pixel, requiring a significantly larger number of transistors per-pixel. Thus, while the pixel size of traditional CMOS image sensors have approached close to 1 micron with a fill factor greater than 90%, current generation ToF sensors can only achieve pixel sizes closer to 10 microns with fill factors closer to 10% [16, 17].

As a consequence ToF sensors with a given footprint (which is typically constrained by die size in the semiconductor fabrication process) will always remain significantly lower resolution than their RGB imaging counterparts. Increasing the overall sensor size (or die size) is generally cost prohibitive as manufacturing cost grows exponentially with the size of the wafer. Therefore, improving ToF spatial resolution without increasing sensor size is an area of significant potential interest [18].

Previously, hybrid ToF systems that combine ToF with other imaging modalities like stereo [19,20], photometric stereo [21], and polarization [22], have been used to achieve super-resolution (SR) performance with commercial ToF cameras. However, these hybrid ToF systems need advanced fusion algorithms and careful registrations between ToF camera and the other imaging modality. Xiao et al. [23] also used deblurring techniques for super resolution using purely software-based techniques. Super-resolution (SR) algorithms were also used to captured ToF images to improve both lateral and depth resolution [18]. Learning-based approaches such as dictionary learning [24, 25] and deep learning [26] have also been used to improve resolution. However, there is a critical difference between our CS-ToF and conventional SR algorithms. Software-based techniques cannot arbitrarily increase resolution. In contrast, CS-ToF perform time-multiplexed optical coding whereby each additional acquired image introduces new spatial information. If temporal resolution is sacrificed, CS-ToF can achieve the full spatial resolution of a DMD. Resolutions as high as 2MP can be achieved using currently available off-the-shelf commercial products.

Can optical multiplexing and compressive sensing (CS) help? Optical multiplexing leverages spatial light modulators (SLMs) to achieve high-resolution imaging with a limited number of sensing elements. Digital micro-mirror devices (DMDs) and liquid crystal on silicon (LCoS) are examples of relatively low cost, commercially available SLMs with at least 1-megapixel resolution. Applications of optical multiplexing include infra-red imaging [27], spectroscopy [28–31], and light transport [32]. By combining compressive sensing and optical multiplexing, we can greatly reduce the number of measurements needed. One of the earliest examples of this is the single pixel camera [33], in which only a single photodiode is used to recover images consisting of 256×256 pixels. A single photodiode with compressive sensing has also been used for depth estimation of the imaged scene [3, 34, 35]. In this paper, we introduce a focal array ToF sensor to increase the measurement bandwidth compared to previous methods using single photodiodes. Compressive imaging models for single-pixel video sensing [36, 37] have been proposed, and recently there has been work on improving the video rate of compressive imaging using 1D line sensors [38] and 2D focal-plane arrays [39]. The key idea in these approaches is to use a spatial light modulation (SLM) to multiplex spatial information into just a few measurements and utilize transform-domain sparsity to reconstruct images at higher resolution than the sensor can natively support.

In this paper, we extend the idea of multiplexed/compressive sensing of spatial information to both intensity and depth images. While the transform-domain sparsity of natural images applies equally well to depth images, the depth is non-linearly related to the intensity measured at each pixel on a ToF sensor. While this property can significantly complicate the reconstruction process, the complication can be avoided by adopting a slightly modified signal representation. Gupta [40] and O'Toole [41] first used a phasor representation to modeling ToF sensors. In the phasor representation, the multiplexing of multiple scene points onto a single sensor measurement can be written as a linear mixing model, thereby allowing us to naturally extend CS-based reconstruction techniques to ToF sensors (CS-ToF).

We present CS-ToF, a novel imaging architecture, to improve the spatial resolution of ToF sensors by performing spatial multiplexing and compressive sensing. We utilize a phasor representation to model the phase and amplitude component of captured correlation signals, resulting in a linear forward model. During CS-based reconstruction, we regularize the amplitude of the reconstructed phasor using a transform-domain sparsity prior. This results in a significant reduction in the number of measurements required for recovery of depth and intensity images with high resolution. In this paper, we introduce the proposed ToF imaging architecture, introduce reconstruction algorithms, and demonstrate a working prototype capable of high-resolution compressive ToF imaging using the proposed framework.

2. CS-ToF

In this section, we first describe our proposed CS-ToF system architecture. We then introduce the ToF imaging model in the presence of the DMD, and explain the DMD-based linear multiplexed measurement model using a phasor representation. Finally, we describe the reconstruction procedure.

2.1. System architecture

The system architecture is shown in Fig. 1. A near-IR laser diode is used to illuminate the scene. The scene is formed on a DMD using an objective lens. The high-resolution DMD-modulated image is then relayed to the low-resolution ToF camera. By changing the coding on the DMD over the course of multiple exposures, we are able to perform spatiotemporal multiplexing of the scene. We can then reconstruct high-resolution amplitude and depth images from multiple low-resolution ToF measurements.

2.2. ToF imaging

ToF is an active 3D imaging technique with a light source as shown in Fig. 2. Both the illumination source (laser diodes) and the shutter of a ToF camera are amplitude-modulated, typically at the

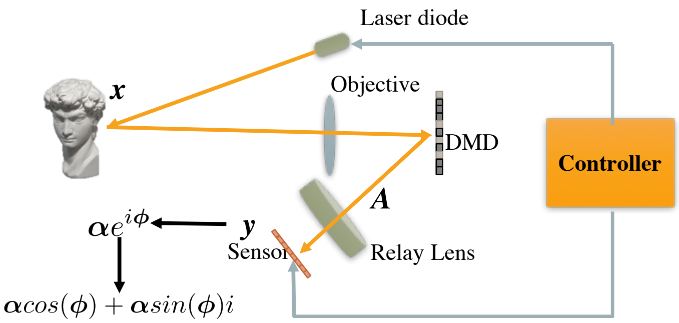
Laser diode
$\boldsymbol{x}$
Objective
DMD
Controller
$\boldsymbol{A}$
$\boldsymbol{\alpha} e^{i\boldsymbol{\phi}}$
$\boldsymbol{y}$
Sensor
Relay Lens
$\boldsymbol{\alpha} cos(\boldsymbol{\phi}) + \boldsymbol{\alpha} sin(\boldsymbol{\phi}) i$

Fig. 1. CS-ToF architecture: The light from the laser diode hits the object and is reflected and imaged on the DMD. Then, the DMD-modulated image is re-imaged at the ToF sensor plane via a relay lens. The laser diode, DMD, and ToF camera are controlled and synchronized by a computer.

same frequency ω. Let the output of the laser diodes be $m(t)$ and the coding at the shutter be $r(t-\psi)$, where ψ is an arbitrary phase delay that can be introduced at the shutter. While the modulated light $m(t)$ travels through space, some part of this light can be reflected by an object at a distance d. Some of this reflected light will reach a sensor pixel p. The light received at the sensor pixel will retain the amplitude modulation frequency ω but will be phase delayed ($\phi_p = \frac{\omega d_p}{2c}$, d_p is the distance of the object) and attenuated ($a_p m(t-\phi_p)$). The sensor measurement at the pixel p, for an exposure duration T will be:

$$B(p,\psi) = \int_{t=0}^{T} a_p m(t-\phi_p) r(t-\psi) dt \tag{1}$$

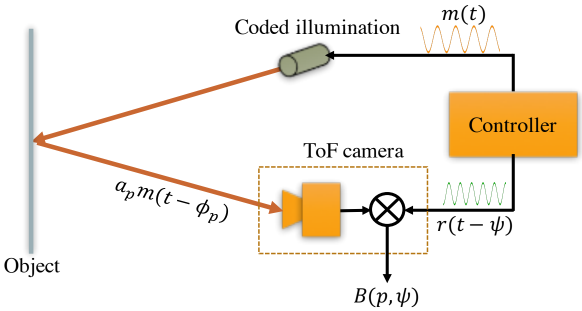

Fig. 2. ToF depth imaging (assume single depth): The computer sends out two signals: $m(t)$ to control the laser diode and $r(t-\psi)$ as reference to ToF sensor. The reflection from object ($a_p m(t-\phi_p)$) is collected by ToF pixels, and then correlates with the reference signal ($r(t-\psi)$) to generate the camera's output.

In most commercial ToF cameras, including the one employed in this paper, the illumination and the reflected signals are of the form:

$$m(t) = o_m + a_m \cos(\omega t) \tag{2}$$

$$r(t) = o_r + a_r \cos(\omega t - \psi) \tag{3}$$

where o_m, a_m, o_r, and a_r are constants. By varying the delay ψ on $r(t)$, one can capture the entire correlation between the reflected signal and the exposure signal. Using demodulation techniques, the reflected signal can be completely recovered. However, most conventional ToF sensors use only four measurements (referred as quadrature measurements) that correspond to $\psi = 0, \pi/2, \pi, 3\pi/2$, to recover the amplitude a_p and the phase ϕ_p of the reflected signal, as given by the formulae:

$$a(p) = \sqrt{\frac{[B(p, 3\pi/2) - B(p, \pi/2)]^2 + [B(p, \pi) - B(p, 0)]^2}{2}} \tag{4}$$

$$\phi(p) = \arctan\left(\frac{B(p, 3\pi/2) - B(p, \pi/2)}{B(p, \pi) - B(p, 0)}\right). \tag{5}$$

Clearly, the phase and amplitude are non-linearly related to the correlational measurements.

2.3. Phasor representation

A linear model relating the scene to the ToF camera measurement is required to recover a high resolution estimate of the scene via compressive sensing. For example, assume two ToF pixels p_1 and p_2 with corresponding amplitude and phase of (a_{p_1}, ϕ_{p_1}) and (a_{p_2}, ϕ_{p_2}). If we combine p_1 and p_2 to form a super-pixel p, the resulting amplitude and the phase at the super-pixel is not $(a_{p_1} + a_{p_2}, \phi_{p_1} + \phi_{p_2})$.

Following the approach of O'Toole [41] and Gupta [40], we use a phasor representation for the ToF output as a complex signal $ae^{i\phi}$ to build a linear model for our system. For consistency, we similarly represent the scene's projection onto the DMD x as a complex value encoding its intensity a_s and phase ϕ_s. The phasor representation for the object's projection on the DMD and ToF sensor are:

$$x = a_s e^{i\phi_s} \tag{6}$$

$$y = ae^{i\phi} \tag{7}$$

Now, we can use this phasor representation to build the linear measurement model from the scene's projection onto the DMD to ToF sensor.

2.4. Measurement model

As shown in Fig. 1, the scene ($\mathbf{x}$) is first projected onto the DMD plane, and modulated with a coded spatial pattern displayed on the DMD. Then, the image on the DMD plane is projected to the ToF camera via the relay lens. We can then write our measurement model as:

$$\mathbf{y} = \mathbf{CMx} = \mathbf{Ax} \tag{8}$$

where $\mathbf{C}$ is the mapping from the DMD pixels to the ToF pixels. $\mathbf{M}$ is the modulation pattern displayed on the DMD. $\mathbf{A} = \mathbf{CM}$, which represents the translation matrix from the scene's projection on DMD to the ToF camera.

The measurement model can be explicitly written as

$$\mathbf{y} = \mathbf{Ax} \;\Rightarrow\; \begin{bmatrix} \alpha_1' e^{i\phi_1'} \\ \vdots \\ \alpha_M' e^{i\phi_M'} \end{bmatrix} = \begin{bmatrix} C_{11} & \cdots & C_{1N} \\ \vdots & \ddots & \vdots \\ C_{M1} & \cdots & C_{MN} \end{bmatrix} \begin{bmatrix} M_1 \\ \vdots \\ \vdots \\ M_N \end{bmatrix} \mathbf{I} \begin{bmatrix} \alpha_1 e^{i\phi_1} \\ \vdots \\ \vdots \\ \alpha_N e^{i\phi_N} \end{bmatrix} \tag{9}$$

where M and N are the total number of ToF pixels and DMD pixels respectively.

During the measurement, we record $\mathbf{y}$ of a given scene $\mathbf{x}$ by T times with displaying a series of patterns on DMD. Assuming the scene $\mathbf{x}$ stays relatively still across the period of T measurements, we can approximate the measurement process as

$$\begin{bmatrix}\mathbf{y_1}\\ \mathbf{y_2}\\ \vdots\\ \mathbf{y_T}\end{bmatrix} = \begin{bmatrix}\mathbf{A_1 x}\\ \mathbf{A_2\ x}\\ \vdots\\ \mathbf{A_T\ x}\end{bmatrix} = \begin{bmatrix}\mathbf{A_1}\\ \mathbf{A_2}\\ \vdots\\ \mathbf{A_T}\end{bmatrix}\mathbf{x} \tag{10}$$

where $\mathbf{A}_t = \mathbf{CM}_t$, $t \in [1, 2, ...T]$. $\mathbf{M}_t$ is the coded pattern displayed on the DMD at time t.

2.5. Reconstruction procedures

From the measurements $\mathbf{y}$ and system matrix $\mathbf{A}$, we will reconstruct the scene's projection on the DMD $\mathbf{x}$. Given the fact that natural images have sparse gradients, we can formulate the reconstruction procedure as the following optimization problem.

$$\widehat{\mathbf{x}} = \underset{\mathbf{x}}{\mathrm{argmin}}\ \frac{1}{2}\|\mathbf{y} - \mathbf{Ax}\|^{\mathbf{2}} + \lambda\mathbf{\Phi}(\mathbf{x}) \tag{11}$$

where λ is a regularization parameter and $\Phi(\mathbf{x})$ is the regularizer.

In this paper, we utilize total variation (TV) as the regularization function defined as

$$\Phi(\mathbf{x}) = \mathrm{TV}(\mathbf{x}) = \sum_i \sqrt{|(G_u(x_i)|^2 + |G_v(x_i)|^2} \tag{12}$$

where $|G_u(x_i)|^2$ and $|G_v(x_i)|^2$ are the horizontal and vertical gradients of 2D image $\mathbf{x}$ at pixel location i. In our experiment, we use TwIST solver to reconstruct the image [42] with more details below.

3. Implementation details

In this section, we discuss some key aspects of our prototype implementation. The hardware prototype is shown in Fig. 3. We use a Fujinon 12.5mm C-Mount Lens to image the scene onto the 1140 × 912-pixel DMD (DLP LightCrafter 4500, Texas Instruments). The DMD-modulated images are then re-imaged using an Edmunds Optics Relay Lens. Finally, a 320 × 240-pixel ToF sensor (OPT8241, Texas Instruments) is placed at the focal plane of the relay lens. During the experiment, imaging areas of 186 × 200 pixels on ToF sensor are used.

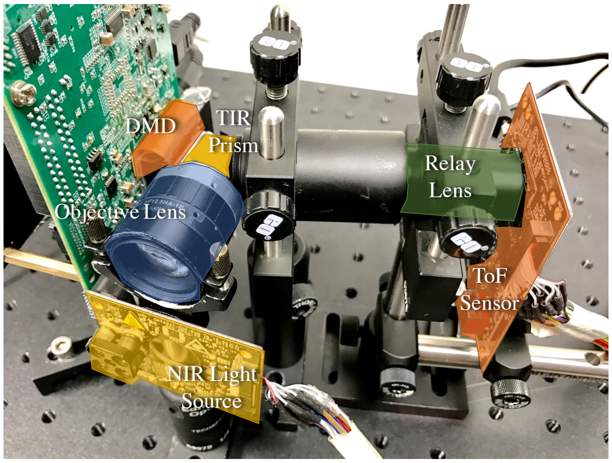

Fig. 3. CS-ToF prototype system with components of the system are highlighted.

3.1. System calibration

The purpose of system calibration is to estimate the system matrix **A**, which depends upon various factors such as the DMD pattern, up-sampling factor, artifacts, as well as optical aberrations and distortion.

As discussed previously, in our system matrix **A** = **CM**, DMD mask **M** is the known pattern displayed on the DMD. Therefore, we need to determine the matrix **C** describing the exact mapping from DMD pixels to ToF camera pixels.

As shown in Fig. 4, first, we display an array of pixel impulses on the DMD, and record the point spread function (PSF) on the ToF camera. We carefully choose the spacing between the DMD impulses to accommodate ToF sensor size and avoid overlapping the PSF on the ToF sensor. As a result, there are 360 impulses per frame. Once we record the image containing the 360 PSFs on the ToF, we select a 5 × 5 neighborhood around each PSF center, and create 360 images only containing one PSF for each image. We vectorize each single-PSF image and insert it into its corresponding column in **C**.

We repeat the above procedures by shifting the impulse array by one pixel, until we traverse every DMD pixel. Eventually, we get a sparse matrix **C** that represents pixel-to-pixel mapping between the DMD and the sensor.

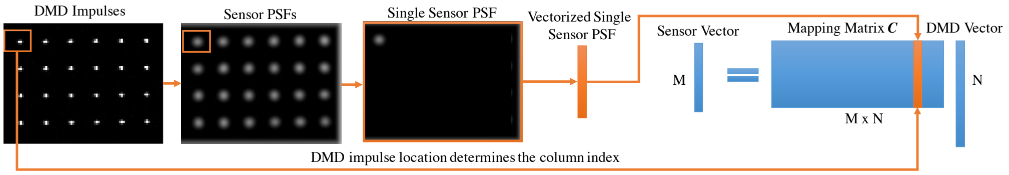

Fig. 4. System Calibration: Calibration is performed by displaying an array of impulses on the DMD and measuring the sensor response for each individual impulse in the array. The response is then placed in the corresponding location in **C**. We traverse every DMD-sensor pixel pair to complete the matrix **C**.

3.2. Modulation pattern on DMD

To ensure measurement quality in the presence of noise, we use Hadamard patterns [31] as the modulation masks displayed on the DMD. We generated a 256 × 256 Hadamard matrix. We use each column of the Hadamard matrix to form a 16 × 16 local pattern. We repeat each local pattern across both the horizontal and vertical directions until it fills the entire DMD plane. We repeat this process to generate all 256 patterns used in our experiments.

3.3. Reconstruction

We use the MATLAB implementation of the TwIST solver [42] to reconstruct the images from the multiplexed compressive measurements. We performed the reconstruction tasks on a Desktop Windows PC with Intel i7 CPU and 32GB RAM running MATLAB with no parallel computing optimization. Reconstructing each output image takes about 30-120 seconds, depending on the compression ratios. With more compression (less number of multiplexed measurements), the reconstruction is faster. The wall time to reconstruct the intensity and depth images with 0.25 compression ratio showing below is about two minutes. The DMD DLP4500 used in this paper does not have a well-designed application programming interface (API) for modulation pattern transmission and synchronizing with the ToF camera. Therefore, extra time is spent on file input/output, synchronization, and customed controlling codes. Potentially, with a better DMD that has good interfacing with the camera and the computer, the wall time can be greatly reduced.

Though we cannot perform real-time video-rate reconstruction currently with TwIST, we do not think the solver would be a potential roadblocks in the future. There are a variety of TV

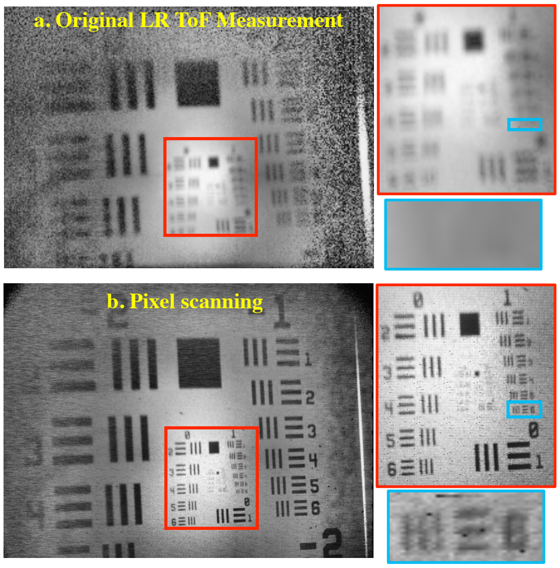

Fig. 5. Pixel scanning of a resolution target: **(a)** shows original low-resolution ToF measurement of the resolution chart target. **(b)** shows the pixel-wise scanning for the resolution target. Color boxes mark corresponding insets.

based complex numerical solver for linear inverse problem available off-the-shelf. One can also exploit the sparsity in transform domain of natural images, such as DCT or wavelet, and use L1/Lasso-based regularizer or solver. If real-time reconstruction is a hard constraint, one can use block-wise parallel reconstruction to accelerate the de-multiplexing. Further more, there are also suitable solvers with GPU acceleration, such as [43].

4. Experiments and results

To demonstrate the performance of our proposed setup, three experiments with resolution chart, Siemens Star, and natural static scene are performed using CS-ToF camera.

4.1. Pixel scanning

Setup: To understand the maximum spatial resolution of the proposed CS-ToF prototype system, we first perform a per-pixel scanning experiment on a USAF 1951 target. In this experiment, we do not acquire multiplexed measurements. Instead, each time, we turn on a DMD pixel and record the response on the ToF sensor. We repeat this process until we scan through all possible DMD pixels. This brute-force process is similar to the one we used for system calibration, except the flat field is replaced by the USAF target. Once finished, an image at the DMD's native resolution is formed.

Pixel scanning reconstruction: Since the resolution target is flat, we receive a flat phase map. Therefore we show only the amplitude pixel-scanning image in Fig. 5. As one can observe, the quality of high-resolution (HR) pixel-scanning results (Fig. 5(b)) is dramatically improved over the original low-resolution (LR) ToF measurement (Fig. 5(a)). We show details of $Group1$ $Element6$ (marked with blue box) in the insets, where details are well-preserved in the pixel-scanning results, but totally missed in the original LR measurement. The pixel scanning experiment result has demonstrated the CS-ToF ability to increase the spatial resolution of the ToF sensor by about 4×.

4.2. Resolution chart

Setup: To evaluate the spatial resolution our CS-ToF prototype can achieve, we perform experiments on standard resolution targets, including the USAF 1951 Resolution Chart and Siemens Star. The size of the resolution chart and Siemens star are approximately $18 \times 15\ cm^2$ and $15 \times 15\ cm^2$, respectively. The target is approximately 0.5 meters away from camera. The experiment setup is visualized in Fig. 6(a). We perform compressive sensing and reconstruction using 4.5 (no compression), 0.6, and 0.25 compression ratios.

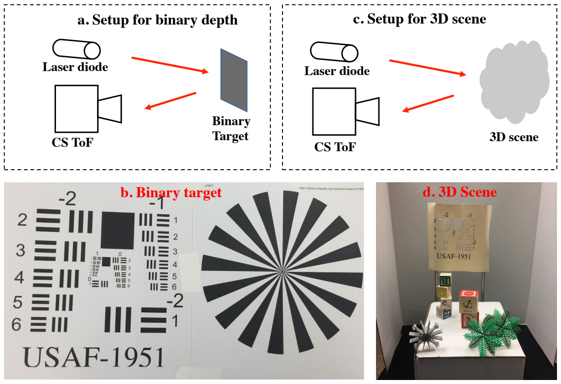

Fig. 6. Real-world experiment setups: **(a)** Conceptual diagram of the resolution target experiment. **(b)** Photo of the resolution targets used. **(c)** Conceptual diagram of the natural scene experiment. **(d)** Photo of the natural scene

Intensity reconstruction: The original LR ToF intensity image and HR intensity images recovered by CS-ToF are shown in Fig. 7. Overall, less compression helps improve the reconstruction quality, but 0.25 compression ratio still provides a qualitatively acceptable reconstruction result.

For the USAF target, we are able to see much finer bars in the HR intensity images recovered by CS-ToF in Figs. 7(b)-7(d), compared to the original LR ToF measurement shown in Fig. 7(a). Particularly, we can see the Element 1 in Group 0 inside the blue bounding box for all CS-ToF results at different compression ratios, which are completely indistinguishable in the original LR measurement. This implies that the resolution improvement is 2 to 3×, which is consistent with the pixel-scanning result.

For the Siemens Star, the original LR ToF measurement fails to characterize the high frequency component close to the center of the star (marked with red box). Whereas the CS-ToF results at different compression ratios are able to resolve the high frequency component.

4.3. 3D natural scene

Setup: To evaluate the real-world performance of the CS-ToF prototype, we perform an experiment on a natural scene. As shown in Figs. 6(c) and 6(d), we construct the scene containing a toy tree, a metal star, two groups of toy bricks, a hollow resolution chart, and a white board, all of which are placed at different depths ranging from 0.5 to $1m$ away from the sensor. We perform compressive sensing and reconstruction using 4.5, 0.6, and 0.25 compression ratios.

Phase or depth reconstruction: The original LR ToF measurement and HR reconstructed phase images using CS-ToF are shown in Fig. 8. Similar to resolution chart results, reconstruction with 0.25 compression ratio can generate a reasonable phase image.

Compared to LR TOF phase image (Fig. 8(a)), more details in the recovered HR phase images (Figs. 8(b)–8(d)) are resolved. For instance, tree leaves (insets marked with red box) can be clearly visualized in the recovered HR phase images with different compression ratios (Figs.

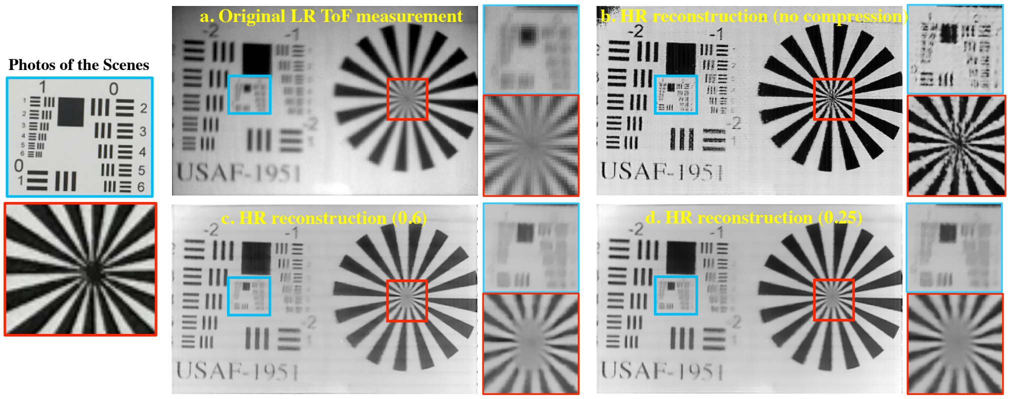
Photos of the Scenes
a. Original LR ToF measurement
b. HR reconstruction (no compression)
c. HR reconstruction (0.6)
d. HR reconstruction (0.25)
USAF-1951

Fig. 7. Intensity reconstruction of resolution charts: **(a)**, **(b)**, **(c)**, and **(d)** show the original LR ToF intensity image, HR CS-ToF reconstruction results with no compression, 0.6 and 0.25 compression ratios, respectively. Fine patterns on resolution chart and the center of Siemens Star are shown in the insets. Ground truth intensity of the insets, taken with a 12-MP camera, are displayed on the left.

8(b2)–8(d2)), but they are obscured in the LR ToF phase image (Fig. 8(a2)). Furthermore, details of a single leaf, in the inset marked with dash black box (Fig. 8(a3)), can be clearly seen in the recovered HR phase images (Figs. 8(b3)–8(d3)).

Another example can be seen from the far “resolution chart". As marked with green box in the photograph in the left, the scene consists of two components at different depths: a portion of a resolution chart with original bars removed and the white board behind. The LR ToF phase image (Fig. 8(a1)) can hardly differentiate the depths in this region. However, the boundary between the two components (marked with red arrows) are preserved and can be clearly seen from the recovered HR phase images (Figs. 8(b1)–8(d1)).

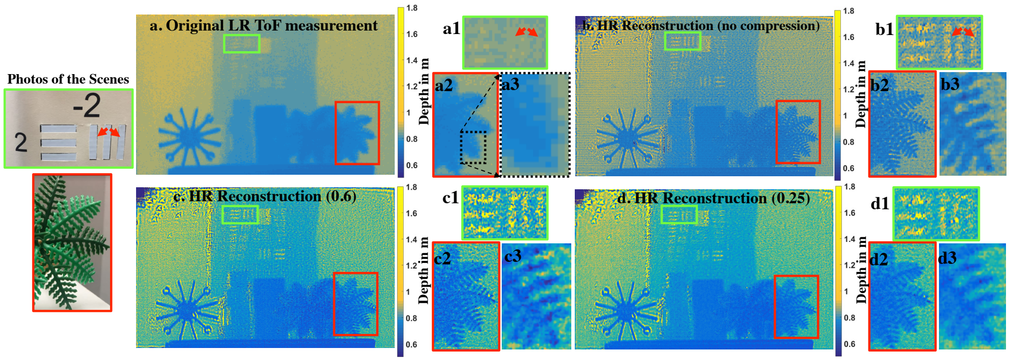

Fig. 8. Phase reconstruction for a natural scene: **(a)**, **(b)**, **(c)**, and **(d)** show LR ToF phase image and HR CS-ToF reconstruction phase images using no compression, 0.6 and 0.25 compression ratios, respectively. Colorbars show the depth information with unit of *meter*. A portion of the far resolution chart and the white board behind is also shown in the insets **(a1-d1)** with their corresponding photograph (marked with green box) in the left. Red arrows point out the boundary of two planes at different depths in **(a1)** and the corresponding photograph. Leaves and their branches on ”toy tree" are shown in the insets **(a2-d2)** with their corresponding photograph (marked with red box) in the left. Close-up images of **(a2-d2)** are further shown in **(a3-d3)**.

Intensity reconstruction: The LR ToF intensity image and recovered HR intensity images are shown in Fig. 9. As we can see, branches of leaves (insets marked with red box) can be seen in the recovered HR intensity images Figs. 9(b2-d2), but are hard to be distinguished in the LR ToF intensity image (Fig. 9(a2)). Other examples can be seen from the center of the metal star: even the screw (marked with red dash circle in Fig. 9(a1)) can be visualized in the recovered HR images (Figs. 9(b1-d1)). More tiny features such as the tip of the spike (red arrows in Fig. 9(a1)) can be observed from the recovered HR intensity images (Figs. 9(b1)–9(d1)).

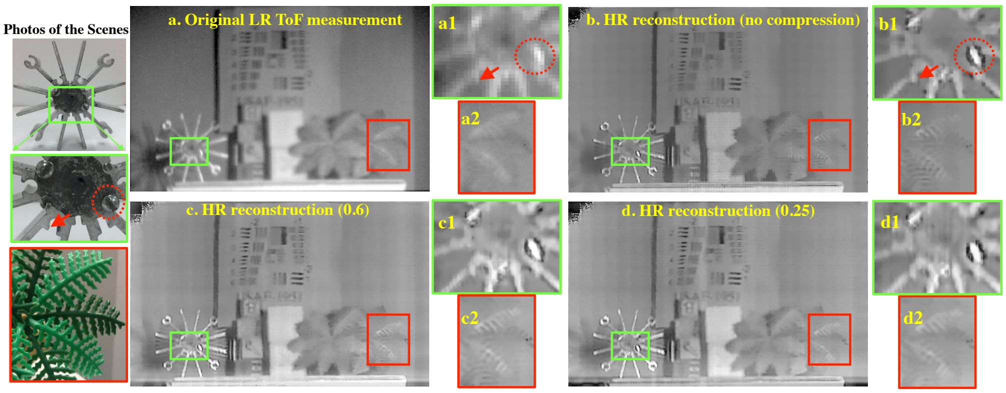
Photos of the Scenes
a. Original LR ToF measurement
a1
a2
b. HR reconstruction (no compression)
b1
b2
c. HR reconstruction (0.6)
c1
c2
d. HR reconstruction (0.25)
d1
d2

Fig. 9. Intensity reconstruction for a natural scene: **(a)**, **(b)**, **(c)**, and **(d)** show LR ToF intensity image and HR CS-ToF reconstruction intensity image using no compression, 0.6 and 0.25 compression ratios, respectively. Fine patterns on the toy tree and the metal star are shown in the insets (**(a1-d1, a2-d2)**) with their corresponding photographs on the left (marked with the green box, and the red box). Note the screw on the metal star (marked with the red dashed circle) and the tip of the metal star (marked with the red arrow).

5. Discussion

Artifacts: Some artifacts can be seen in the recovered intensity images. The artifact is due to imperfect alignments and calibration for the **A** matrix. This can be minimized by more careful calibration or advanced noise subtraction algorithms. In this paper, we tried background subtraction, non-local means filter and band-pass filter in the Fourier domain to minimize the artifacts in the recovered intensity images.

Small ToF sensor fill factor: In our proposed setup, multiple (eg. m pixels) DMD pixels approximately project onto one pixel of the ToF sensor. Theoretically, the scene on the DMD should be a uniform brightness with darker at the periphery due to vignetting, unlike our observation in Fig. 10(a). This discrepancy is likely due to the low fill factor of the ToF, which causes missed DMD-ToF mapping information to be missed in the calibration of matrix **A**. This in turn causes aliasing, visible in Fig. 10(b), an effect typical of all low fill factor sensors. We can mitigate this effect with conventional sensor anti-aliasing approaches, such as placing a diffuser on the surface of the sensor or slightly defocusing the projection onto the ToF camera. The effect of such low-pass filtering can be seen in Figs. 10(c)–10(d).

Compressive reconstruction of complex value: Compressive reconstruction of complex value is an interesting challenge [44]. Methods have been discussed for different imaging models using CS reconstruction of complex inputs such as terahertz (THz) imaging, synthetic aperture radar, holography, etc [45,46]. Regularizer and parameter choice can have significant effect the reconstruction quality, including CS reconstructions using our proposed architecture. Future work may provide more insight into the best selection of CS approaches for this particular application.

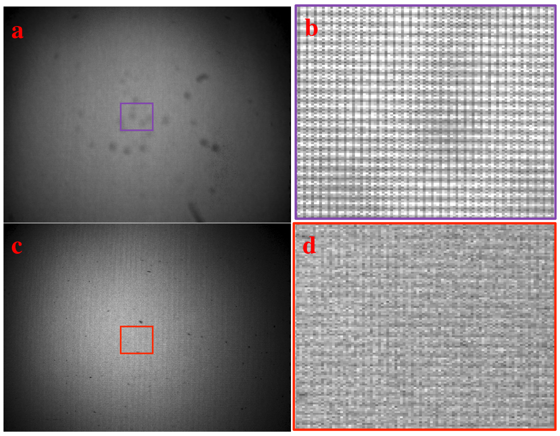

Fig. 10. Scene projected on DMD plane with white field illumination: **(a)**. The scene on DMD with ToF camera placed at the back focal plane of the relay lens. **(c)**. The same scene on DMD with ToF camera slightly defocused. Color boxes represent insets in (a) and (c).

Spatiotemporal tradeoff: A key advantage of the CS-ToF is the flexibility of trading-off among spatial resolution, temporal resolution and image quality. The maximum spatial resolution (g) is limited by the physical resolution of the DMD or SLM, which is $g = 1.04$ megapixel (MP) (1140 × 912) in our prototype. The ToF sensor in our prototype has a usable imaging area of $s = 186 \times 200 = 0.037$MP and can operate at $f = 150 fps$ maximum. Therefore the maximum measurement bandwidth $b = f \cdot s = 5.55\text{MP}/s$. The temporal resolution (t), and image quality is dependent on the number of measurement M we use for reconstruction. At each measurement, we take s coded samples of the "ground truth" image on the DMD. For example, if image quality is not a concern, we can use $M = 1$ measurement to perform the reconstruction, therefore $c = s \cdot m/g = 3.6\%$ compression ratio is achieved, and the temporal resolution is $t = f/m = 150 fps$. As demonstrated in our manuscript, high-quality reconstruction requires a minimum of $M = 7$ frames, resulting $0.037 \times 7/1.04 = 0.25$ compression ratio and $150/7 = 21.4 fps$ temporal resolution.

Temporal (depth) super resolution: The phasor representation can be a linear model in spatial domain for ToF, but it is non-linear in the temporal domain which prevents us from using our proposed setup for depth super resolution.

Depth accuracy: A simulation experiment has been performed to quantify the depth accuracy of our CS-ToF framework. We assume that DMD has 1140 × 912 pixels and the ToF sensor has 120 × 153 pixels in the simulation experiment. A 3D scene (Fig. 11(a)) with ground truth depth (Fig. 11(b)) is chosen from the Middlebury Dataset [47]. The intensity and depth images of the ground truth scene have the size of 1140 × 912 pixels. The translation matrix from the DMD plane to the ToF sensor plane in our CS-ToF framework is simulated in the same method described in Section 3.1 with a custom-defined PSF. The responses on the ToF sensor are acquired using the forward model described above with Hadamard patterns used in Section 3.2. We then reconstruct the HR CS-ToF images with the same reconstruction algorithm described in Section 2.5 using 0.6 and 0.2 compression ratios. Gaussian noises with signal-to-noise ratios (SNR) of 30dB, 25dB, and 20dB are added in the ToF measurements.

To quantify the depth accuracy of our CS-ToF, depth values from the same regions marked with red lines in the ground truth depth image (Fig. 11(b)) and the HR CS-ToF reconstruction depth image with 0.6 compression ratio (Fig. 11(d)) and 0.2 compression ratio (Fig. 11(e)), are selected and compared. To make a fair comparison, the bicubic interpolation of LR ToF measurement depth (Fig. 11(c)) is also generated by down-sampling the ground truth image to 120 × 153 pixels as the regular ToF response and then up-sampling to the same size with the ground truth. Figures 11(c)–11(e) are generated with 25 dB SNR Gaussian noise in the measurements. Figures 11(f)–11(h) show the depth values of pixels along the red lines with different Gaussian noises added into the measurements. The root mean square error (RMSE) of HR CS-ToF reconstruction depth compared to the ground truth depth is then calculated using the data shown in Figs. 11(d) and 11(e). We also quantify the RMSE of LR ToF depth with bicubic interpolation compared to the ground truth depth. The results are summarized in Table 1. Although the depth accuracy of CS-ToF might be worse compared to the regular ToF imaging due to optical multiplexing in the CS-ToF, it has better depth accuracy compared to that of bicubic interpolation of the LR ToF measurement.

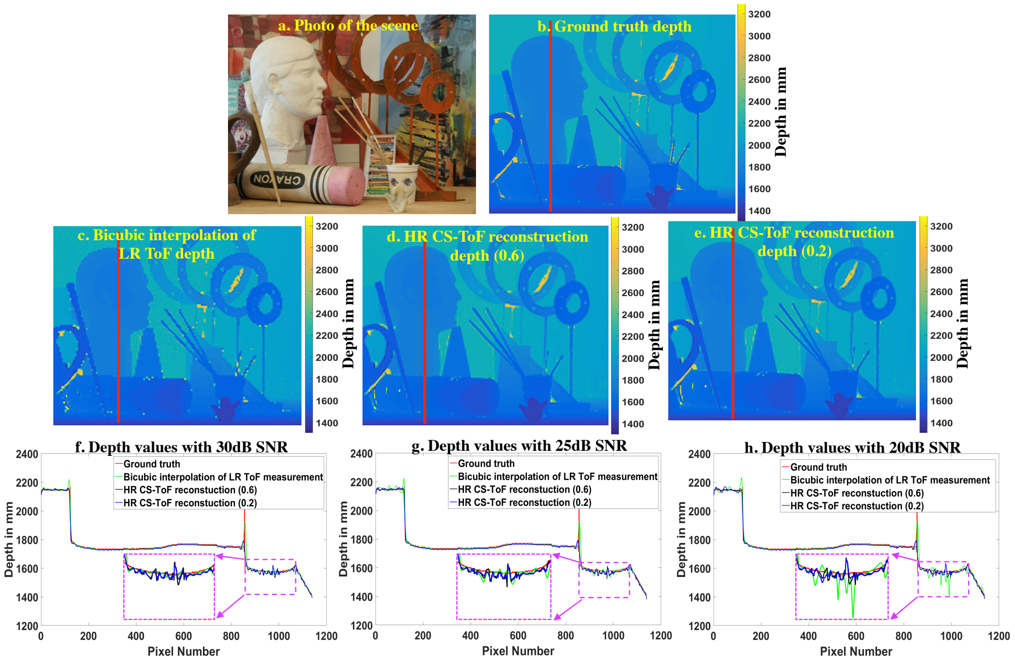

Fig. 11. The quantification of depth accuracy for CS-ToF: **(a)**. The photograph of the 3D scene for the simulation experiments. **(b)**. The ground truth depth for the 3D scene. **(c)**. The bicubic interpolation of LR ToF measurement depth with 25 dB Gaussian noise added in the system. **(d), (e)** show the HR CS-ToF depth iamges with 0.6 and 0.2 compression ratios, respectively. (25 dB Gaussian noise is added in the measurements). **(f)** shows the depth values along the red lines in **(b-e)** with 30dB SNR Gaussian noise in the measurements. **(g)** shows the depth values on the same pixels with **(f)** with 25dB SNR Gaussian noise added. **(h)** shows the depth values on the same pixels with **(f)** with 20dB SNR Gaussian noise added.

Table 1. RMSE of LR ToF measurement depth with bicubic interpolation and HR CS-ToF reconstruction depth with respect to the grouth truth depth.

	RMSE with 30dB SNR (*mm*)	RMSE with 25dB SNR (*mm*)	RMSE with 20dB SNR (*mm*)
Bicubic Interpolation of LR ToF	24.2	25.2	28.3
HR CS-TOF reconstruction (0.6)[1]	19.3	19.3	19.7
HR CS-TOF reconstruction (0.2)[2]	19.0	19.3	19.5

[1] (0.6): The reconstruction is with the compression ratio of 0.6.
[2] (0.2): The reconstruction is with the compression ratio of 0.2.

6. Conclusion

In this paper, we have proposed an architecture for high spatial resolution ToF imaging. We utilize a phasor representation to achieve a linear compressive sensing model, which we demonstrate using experimental hardware. We believe CS-ToF camera has provided a simple and cost-effective solution for SR 3D imaging, which might benefit many 3D imaging applications such as improving the accuracy for 3D detection and tracking.

Funding

National Science Foundation (NSF) (CAREER IIS-1453192, CAREER IIS-1652633, and CCF-1527501); Office of Naval Research (ONR) (N00014-15-1-2735); Defense Advanced Research Projects Agency (DARPA) (REVEAL HR0011-16-C-0028); Texas Instruments (TI) Graduate Student Fellowship.