Exploiting sparsity in time-of-flight range acquisition using a single time-resolved sensor

Ahmed Kirmani, Andrea Colaço, Franco N. C. Wong, and Vivek K. Goyal*

Research Laboratory of Electronics, Massachusetts Institute of Technology,
Cambridge, Massachusetts 02139, USA

**v.goyal@ieee.org*

http://www.rle.mit.edu/stir/

Abstract: Range acquisition systems such as light detection and ranging (LIDAR) and time-of-flight (TOF) cameras operate by measuring the time difference of arrival between a transmitted pulse and the scene reflection. We introduce the design of a range acquisition system for acquiring depth maps of piecewise-planar scenes with high spatial resolution using a single, omnidirectional, time-resolved photodetector and no scanning components. In our experiment, we reconstructed 64×64-pixel depth maps of scenes comprising two to four planar shapes using only 205 spatially-patterned, femtosecond illuminations of the scene. The reconstruction uses parametric signal modeling to recover a set of depths present in the scene. Then, a convex optimization that exploits sparsity of the Laplacian of the depth map of a typical scene determines correspondences between spatial positions and depths. In contrast with 2D laser scanning used in LIDAR systems and low-resolution 2D sensor arrays used in TOF cameras, our experiment demonstrates that it is possible to build a non-scanning range acquisition system with high spatial resolution using only a standard, low-cost photodetector and a spatial light modulator.

OCIS codes: (110.6880) Three-dimensional image acquisition; (110.1758) Computational imaging.

References and links

1. K. Carlsson, P. E. Danielsson, R. Lenz, A. Liljeborg, L. Majlöf, and N. Åslund, “Three-dimensional microscopy using a confocal laser scanning microscope,” Opt. Lett. **10**, 53–55 (1985).
2. J. Sharpe, U. Ahlgren, P. Perry, B. Hill, A. Ross, J. Hecksher-Sørensen, R. Baldock, and D. Davidson, “Optical projection tomography as a tool for 3d microscopy and gene expression studies,” Science **296**, 541–545 (2002).
3. A. Wehr and U. Lohr, “Airborne laser scanning—an introduction and overview,” ISPRS J. Photogramm. Remote Sens. **54**, 68–82 (1999).
4. D. A. Forsyth and J. Ponce, *Computer Vision: A Modern Approach* (Prentice-Hall, 2002).
5. S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, “A comparison and evaluation of multi-view stereo reconstruction algorithms,” in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, (2006), pp. 519–528.
6. S. Hussmann, T. Ringbeck, and B. Hagebeuker, “A performance review of 3D TOF vision systems in comparison to stereo vision systems,” in *Stereo Vision*, A. Bhatti, ed. (InTech, 2008), pp. 103–120.
7. E. Stoykova, A. A. Alatan, P. Benzie, N. Grammalidis, S. Malassiotis, J. Ostermann, S. Piekh, V. Sainov, C. Theobalt, T. Thevar, and X. Zabulis, “3-D time-varying scene capture technologies—A survey,” IEEE Trans. Circuits Syst. Video Technol. **17**, 1568–1586 (2007).

#153103 - $15.00 USD Received 19 Aug 2011; revised 2 Oct 2011; accepted 3 Oct 2011; published 17 Oct 2011

8. D. Scharstein and R. Szeliski, “A taxonomy and evaluation of dense two-frame stereo correspondence algorithms,” Int. J. Comput. Vis. **47**, 7–42 (2002).
9. B. Schwarz, “LIDAR: mapping the world in 3D,” Nat. Photonics **4**, 429–430 (2010).
10. S. B. Gokturk, H. Yalcin, and C. Bamji, “A time-of-flight depth sensor — system description, issues and solutions,” in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops*, (2004), p. 35.
11. S. Foix, G. Alenyà, and C. Torras, “Lock-in time-of-flight (ToF) cameras: a survey,” IEEE Sens. J. **11**, 1917–1926 (2011).
12. A. P. Cracknell and L. W. B. Hayes, *Introduction to Remote Sensing* (Taylor & Francis, 1991).
13. F. Blais, “Review of 20 years of range sensor development,” J. Electron. Imaging **13**, 231–240 (2004).
14. R. Lamb and G. Buller, “Single-pixel imaging using 3d scanning time-of-flight photon counting,” SPIE Newsroom (2010). DOI: 10.1117/2.1201002.002616.
15. A. Medina, F. Gayá, and F. del Pozo, “Compact laser radar and three-dimensional camera,” J. Opt. Soc. Am. A. **23**, 800–805 (2006).
16. S. Schuon, C. Theobalt, J. Davis, and S. Thrun, “High-quality scanning using time-of-flight depth superresolution,” in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops*, (2008).
17. M. Vetterli, P. Marziliano, and T. Blu, “Sampling signals with finite rate of innovation,” IEEE Trans. Signal Process. **50**, 1417–1428 (2002).
18. T. Blu, P.-L. Dragotti, M. Vetterli, P. Marziliano, and L. Coulot, “Sparse sampling of signal innovations,” IEEE Signal Process. Mag. **25**, 31–40 (2008).
19. M. Sarkis and K. Diepold, “Depth map compression via compressed sensing,” in *Proceedings of IEEE International Conference on Image Processing*, (2009), pp. 737–740.
20. I. Tošić, B. A. Olshausen, and B. J. Culpepper, “Learning sparse representations of depth,” IEEE J. Sel. Top. Signal Process. **5**, 941–952 (2011).
21. E. J. Candès and T. Tao, “Near-optimal signal recovery from random projections: Universal encoding strategies?” IEEE Trans. Inform. Theory **52**, 5406–5425 (2006).
22. D. L. Donoho, “Compressed sensing,” IEEE Trans. Inform. Theory **52**, 1289–1306 (2006).
23. M. B. Wakin, J. N. Laska, M. F. Duarte, D. Baron, S. Sarvotham, D. Takhar, K. F. Kelly, and R. G. Baraniuk, “An architecture for compressive imaging,” in *Proceedings of IEEE International Conference on Image Processing*, (2006), pp. 1273–1276.
24. M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. Kelly, and R. G. Baraniuk, “Single-pixel imaging via compressive sampling,” IEEE Signal Process. Mag. **25**, 83–91 (2008).
25. M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing* (Springer, 2010).
26. G. Howland, P. Zerom, R. W. Boyd, and J. C. Howell, “Compressive sensing LIDAR for 3D imaging,” in *CLEO:2011 - Laser Applications to Photonic Applications*, OSA Technical Digest (CD) (Optical Society of America, 2011), paper CMG3.
27. P. L. Dragotti, M. Vetterli, and T. Blu, “Sampling moments and reconstructing signals of finite rate of innovation: Shannon meets Strang-Fix,” IEEE Trans. Signal Process. **55**, 1741–1757 (2007).
28. A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, 3rd ed. (Prentice-Hall, 2009).
29. M. Grant and S. Boyd, “CVX: Matlab software for disciplined convex programming, version 1.21,” http://cvxr.com/cvx.
30. M. Grant and S. Boyd, “Graph implementations for nonsmooth convex programs,” in *Recent Advances in Learning and Control*, V. Blondel, S. Boyd, and H. Kimura, eds. (Springer-Verlag Limited, 2008), pp. 95–110.
31. G. C. M. R. de Prony, “Essai éxperimental et analytique: Sur les lois de la dilatabilité de fluides élastique et sur celles de la force expansive de la vapeur de l’alkool, à différentes températures,” J. de l’École Polytechnique **1**, 24–76 (1795).

1. Introduction

Sensing 3D scene structure is an integral part of applications ranging from 3D microscopy [1,2] to geographical surveying [3]. While 2D imaging is a mature technology, 3D acquisition techniques have room for significant improvements in spatial resolution, range accuracy, and cost effectiveness. Humans use both monocular cues—such as motion parallax—and binocular cues—such as stereo disparity—to perceive depth, but camera-based stereo vision techniques [4] suffer from poor range resolution and high sensitivity to noise [5, 6]. Computer vision techniques—including structured-light scanning, depth-from-focus, depth-from-shape, and depth-from-motion [4, 7, 8]—are computation intensive, and the range output from these methods is highly prone to errors from miscalibration, absence of sufficient scene texture, and

#153103 - $15.00 USD Received 19 Aug 2011; revised 2 Oct 2011; accepted 3 Oct 2011; published 17 Oct 2011

low signal-to-noise ratio (SNR) [5, 6, 8].

In comparison, active range acquisition systems such as LIDAR systems [9] and TOF cameras [10, 11] are more robust against noise [6], work in real-time at video frame rates, and acquire range information from a single viewpoint with little dependence on scene reflectance or texture. Both LIDAR and TOF cameras operate by measuring the time difference of arrival between a transmitted pulse and the scene reflection. LIDAR systems consist of a pulsed illumination source such as a laser, a mechanical 2D laser scanning unit, and a single time-resolved photodetector or avalanche photodiode [9, 12–14]. The TOF camera illumination unit is composed of an array of omnidirectional, modulated, infrared light emitting diodes (LEDs) [10, 11, 15]. The reflected light from the scene—with time delay proportional to distance—is focused at a 2D array of TOF range sensing pixels. A major shortcoming of LIDAR systems and TOF cameras is low spatial resolution, or the inability to resolve sharp spatial features in the scene. For real-time operability LIDAR devices have low 2D scanning resolution. Similarly, due to limitations in the 2D TOF sensor array fabrication process and readout rates, the number of pixels in TOF camera sensors is also currently limited to a maximum of 320×240 pixels [15, 16]. Consequently, it is desirable to develop novel, real-time range sensors that possess high spatial resolution without increasing the device cost and complexity.

1.1. Main contribution

Natural scenes are often primarily constituted of planar facets. In this paper, we introduce a framework for acquiring the depth map of a piecewise-planar scene at high range and spatial resolution using only a single photodetector as the sensing element and a spatiotemporally-modulated light source as the illumination unit. In our framework (see Fig. 1), an omnidirectional, temporally-modulated periodic light source illuminates a spatial light modulator (SLM) with an $N \times N$ pixel resolution, which then projects a chosen 2D spatial pattern on the piecewise-planar scene. The light reflected from the illuminated portions of the scene is then focused at a time-resolving photodetector and digitized into K digital samples by an analog-to-digital converter (ADC) that is synchronized with the light source. This measurement process is repeated M times; depending on the desired spatial resolution, M typically ranges from 1 to 5% of the total number of pixels in the SLM. The recorded time samples are computationally processed to obtain a 2D scene depth map at the same pixel resolution as the SLM.

In our framework, the sequence of SLM configurations and the computational processing each proceed in two steps. Both steps exploit implicit or explicit modeling of the scene as piecewise planar.

Step 1 uses no spatial patterning from the SLM, i.e., a fully-transparent configuration. Under the assumption that the scene is approximately piecewise planar, the continuous-time light intensity signal at the single photodetector is approximated well in a certain parametric class. Estimation of the parameters of the signal implies recovery of the range of depth values present in the scene. Note that the use of a parametric signal modeling and recovery framework [17, 18] enables us to achieve high depth resolution relative to the speed of the time sampling at the photodetector. After discretizing the depths identified in this step, the remaining problem is to find correspondences between spatial locations and depths to form the depth map.

Step 2 uses many pseudorandom binary patterns on the SLM. The assumption that the scene is approximately piecewise planar translates to the Laplacian of the depth map being approximately sparse. We introduce a novel convex optimization problem that finds the depth map consistent with the measurements that approximately minimizes the number of nonzero entries in the Laplacian of the depth map. Solving this optimization problem with a general-purpose software package yields the desired depth map.

#153103 - $15.00 USD Received 19 Aug 2011; revised 2 Oct 2011; accepted 3 Oct 2011; published 17 Oct 2011

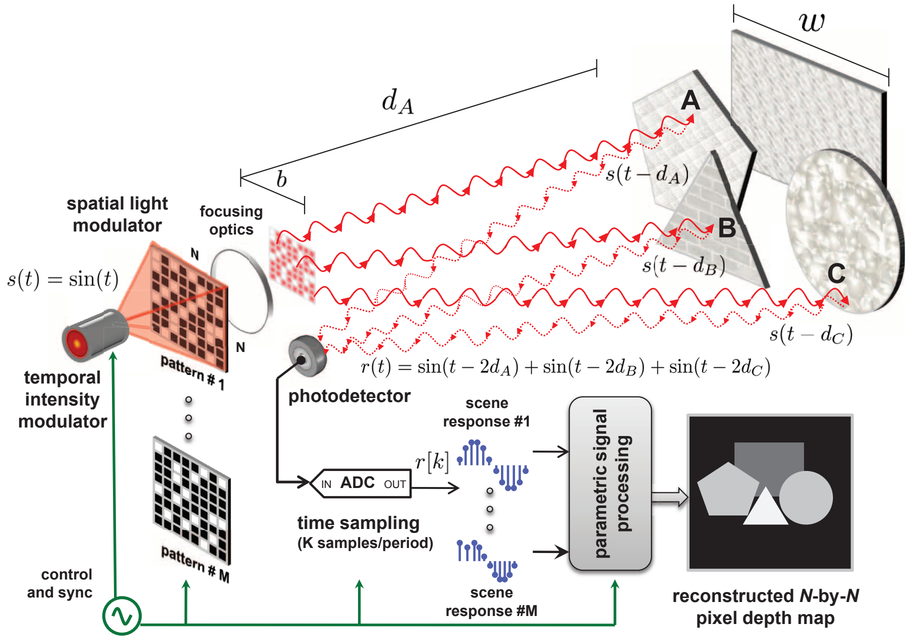

Fig. 1. The proposed architecture for acquiring depth maps of scenes constituted of piecewise-planar facets. The scene is in far field, i.e., the baseline b and the dimensions of each planar facet w are much smaller than the distance between the imaging device and the scene. A light source with periodically-varying intensity $s(t)$ illuminates an $N \times N$-pixel SLM. The scene is serially illuminated with M chosen spatial patterns. For each patterned illumination the reflected light is focused at the photodetector and K digital time samples are recorded. The total $M \times K$ time samples are computationally processed to reconstruct an $N \times N$-pixel depth map of the scene.

1.2. *Related work*

The use of pseudorandom binary SLM configurations and the exploitation of transform-domain sparsity of natural scene depth [19,20] are reminiscent of compressed sensing [21,22] (CS) and, more specifically, the concept of a "single-pixel camera" [23, 24]. CS provides techniques to estimate a signal vector x from linear measurements of the form $y = Ax + w$, where w is additive noise and vector y has *fewer* entries than x. The estimation methods exploit that there is a linear transformation T such that Tx is approximately sparse.

The depth map of a scene is generally more compressible or sparse than the reflectance or texture (see Fig. 2). Thus, we expect a smaller number of measurements to suffice; this is indeed the case, as our number of measurements is 1 to 5% of the number of pixels as compared to 10 to 40% for reflectance imaging [23, 24].

In a preliminary application of the CS framework to LIDAR systems [26], 2 ns square pulses from a function generator drive a 780 nm laser diode to illuminate a scene with spatial patterning provided by a digital micromirror device. Incident reflected light is measured with a photon-counting detector and gated to collect photons arriving from an *a priori* chosen range interval, and then conventional CS reconstruction is applied to recover an image of the objects within the selected depth interval. The use of impulsive illumination and range gating make this a conventional CS problem in that the quantities of interest (reflectances as a function of spatial position, within a depth range) are combined linearly in the measurements. This ap-

#153103 - $15.00 USD Received 19 Aug 2011; revised 2 Oct 2011; accepted 3 Oct 2011; published 17 Oct 2011

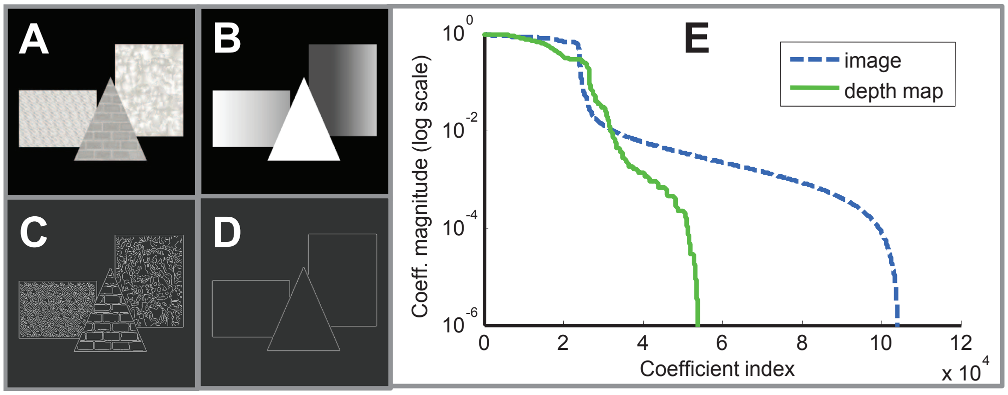

Fig. 2. *Sparsity* of a signal (having a basis expansion or similar representation with a small number of coefficients significantly different from zero) is widely exploited for signal estimation and compression [25]. An $N \times N$-pixel digital photograph (A) or depth map (B) of a scene requires N^2 pixel values for representation in the spatial domain. As illustrated with the output of an edge-detection method, the Laplacian of a depth map (D) typically has fewer significant coefficients than the Laplacian of a photograph (C). This structure of natural scenes is also reflected in discrete wavelet transform (DWT) coefficients sorted by magnitude: a photograph has slower decay of DWT coefficients and more nonzero coefficients (E: blue, dashed) than the corresponding depth map (E: green, solid). We exploit this simplicity of depth maps in our range acquisition framework.

proach achieves 3D imaging with a single sensor, but it has two major disadvantages: acquiring a complete scene depth map requires a full range sweep; and there is no method to distinguish between objects at different depths within a chosen range interval. The proof-of-concept system [26] has 30 cm range resolution and 32 × 32 pixel resolution.

In our framework, depths are revealed through phase offsets between the illumination signal and the reflected light rather than by direct measurement of time delays. Conventional CS reconstruction techniques are inapplicable because the quantities of interest (depths as a function of spatial position) are combined nonlinearly in the measurements. For example, consider three points illuminated with intensity $s(t) = \sin(t)$ as shown in Fig. 1. The sum of the reflected returns has the form $r(t) = a\sin(t-\phi)$, where the amplitude $a = [3+2\cos(2d_A - 2d_B)+2\cos(2d_A - 2d_C)+2\cos(2d_B - 2d_C)]^{1/2}$ and phase shift $\phi = -\tan^{-1}[(\sin(2d_A)+\sin(2d_B)+\sin(2d_C))/(\cos(2d_A)+\cos(2d_B)+\cos(2d_C))]$ nonlinearly combine the depths d_A, d_B, and d_C. The parameters a and ϕ can be estimated using samples of $r(t)$, but the three unknown depths cannot be uniquely determined from a and ϕ, and moreover all spatial resolution is lost due to the omnidirectional collection of light at the photodetector. Varying the SLM configuration would produce different nonlinear mixtures of depths and thus could make the solution unique, but the complexity stemming from nonlinearity of mixing remains. Our approach avoids this "nonlinear CS" formulation by creating intermediate quantities that represent, for each of a set of discrete depths, the sum of reflectances at the given depth.

1.3. Outline

The remainder of the paper is organized as follows: Section 2 establishes notation for our imaging setup. Sections 3 and 4 discuss the modeling and computational recovery associated with Steps 1 and 2, respectively, with the scene restricted to a single planar, rectangular facet for clarity of exposition. Section 5 describes the extensions of the framework that allow us to handle

#153103 - $15.00 USD Received 19 Aug 2011; revised 2 Oct 2011; accepted 3 Oct 2011; published 17 Oct 2011

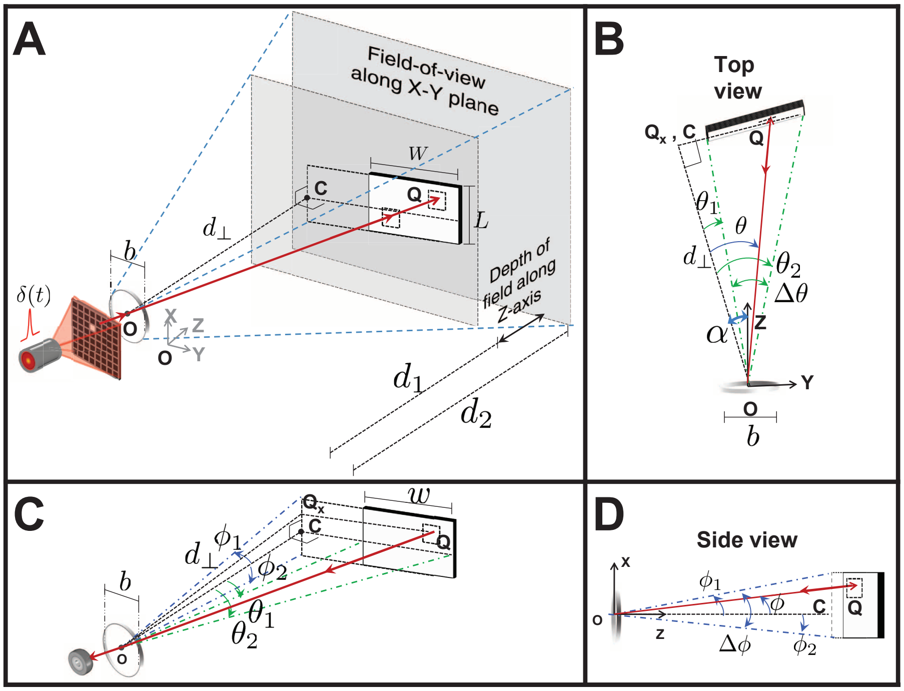

Fig. 3. (A) Scene setup for parametric signal modeling of TOF light transport; (B) Top view; (C) Notation for various angles; (D) Side view.

scenes with multiple planar facets that are not necessarily rectangular. The experiment is described in Section 6, and further extensions to textured scenes and non-impulsive illumination are discussed in Section 7. Section 8 concludes the paper.

2. Notation and assumptions for analysis of a single rectangular facet

Consider the setup shown in Fig. 3. A chosen SLM pattern is focused on the scene using a focusing system as shown in Fig. 3A. The center of the focusing system is denoted by O and is also the origin for a 3D coordinate system (X,Y,Z). All angles and distances are measured with respect to this global coordinate system. The focusing optics for the SLM illumination unit are chosen such that it has a depth-of-field (DOF) between distances d_1 and d_2 ($d_1 < d_2$) along the Z dimension and a square field-of-view (FOV) along the X-Y axes. Thus, the dimensions of a square SLM pixel projected onto the scene remains constant within the DOF and across the FOV. We denote the dimensions of an SLM pixel within the DOF by $\Delta \times \Delta$. An SLM with higher spatial resolution corresponds to a smaller value of Δ. We also assume that the scene lies within the DOF so that all planar facets in the scene are illuminated by projection pixels of the same size. In our mathematical modeling and experiments, we only consider binary patterns, i.e., each SLM pixel is chosen to be either completely opaque or fully transparent. In Section 7, we discuss the possibility of using continuous-valued or gray-scale SLM patterns to compensate for rapidly-varying scene texture and reflectance.

The light reflected from the scene is focused at the photodetector. Note that we assume that the baseline separation b between the focusing optics of the detector and the SLM illumination optics is very small compared to the distance between the imaging device and the scene; i.e.,

Received 19 Aug 2011; revised 2 Oct 2011; accepted 3 Oct 2011; published 17 Oct 2011

if Q is a scene point as shown in Fig. 3, the total path length $O \to Q \to$ photodetector is approximately equal to the path length $O \to Q \to O$. Thus, we may conveniently model O as the effective optical center of the entire imaging setup (illumination and detector).

Sections 3 and 4 provide analyses of the time-varying light intensity at the detector in response to impulse illumination of a scene containing a single rectangular planar facet. The dimensions of the facet are $W \times L$. Let OC be the line that lies in the Y-Z plane and is also perpendicular to the rectangular facet. The plane is tilted from the zero-azimuth axis (marked Z in Fig. 3), but the developments of Section 3 will show that this tilt is immaterial in our approach to depth map construction. For simplicity, we assume no tilt from the zenith axis (marked X in Fig. 3); a nonzero tilt would be immaterial in our approach.

The following parameters completely specify the rectangular facet (see Fig. 3C):

- $d_\perp$ denotes the length of the line OC.
- ϕ_1 and ϕ_2 are angles between line OC and the extreme rays connecting the vertical edges of the rectangular facet to O, and $\Delta\phi = |\phi_1 - \phi_2|$ is their difference; clearly, $\Delta\phi$ is related to L.
- θ_1 and θ_2 are angles between line OC and the extreme rays connecting the horizontal edges of the rectangular facet to O, and $\Delta\theta = |\theta_1 - \theta_2|$ is their difference; clearly, $\delta\theta$ is related to W.
- α is the angle between OC and the Z axis in the Y-Z plane.

For our light transport model, we assume that the scene is in the far field, i.e., the dimensions of the rectangular facet are small compared to the distance between the scene and the imaging device, or $W \ll d_1$ and $L \ll d_1$. This implies that $\Delta\phi$ and $\Delta\theta$ are small angles and that the radial fall-off attenuation of light arriving from different points on the rectangular facet is approximately the same for all the points. For developing the basic light transport model we also assume that the rectangular facet is devoid of texture and reflectance patterns. When a 2D scene photograph or image is available prior to data acquisition, then this assumption can be relaxed without loss of generality as discussed in Section 7. Finally, we set the speed of light to unity so that the numerical value of the time taken by light to traverse a given distance is equal to the numerical value of the distance.

3. Response of a single rectangular facet to fully-transparent SLM pattern

3.1. *Scene response*

Let Q be a point on the rectangular planar facet at an angle of θ ($\theta_1 < \theta < \theta_2$) and ϕ ($\phi_1 < \phi < \phi_2$) with respect to the line OC as shown in Fig. 3. A unit-intensity illumination pulse, $s(t) = \delta(t)$, that originates at the source at time $t = 0$ will be reflected from Q, attenuated due to scattering, and arrive back at the detector delayed in time by an amount proportional to the distance $2|OQ|$. Since the speed of light is set to be unity, the delay is exactly equal to the distance $2|OQ|$. Thus the signal incident on the photodetector in response to impulse illumination of Q is mathematically given by

$$q(t) \ = \ a\,\delta(t - 2|OQ|),$$

where a is the total attenuation (transmissivity) of the unit-intensity pulse. Since the photodetector has an impulse response, denoted by $h(t)$, the electrical output $r_q(t)$ of the photodetector is mathematically equivalent to convolution of the signal $q(t)$ and the detector response $h(t)$:

$$r_q(t) \ = \ h(t) * a\,\delta(t - 2|OQ|) \ = \ a\,h(t - 2|OQ|).$$

#153103 - $15.00 USD Received 19 Aug 2011; revised 2 Oct 2011; accepted 3 Oct 2011; published 17 Oct 2011

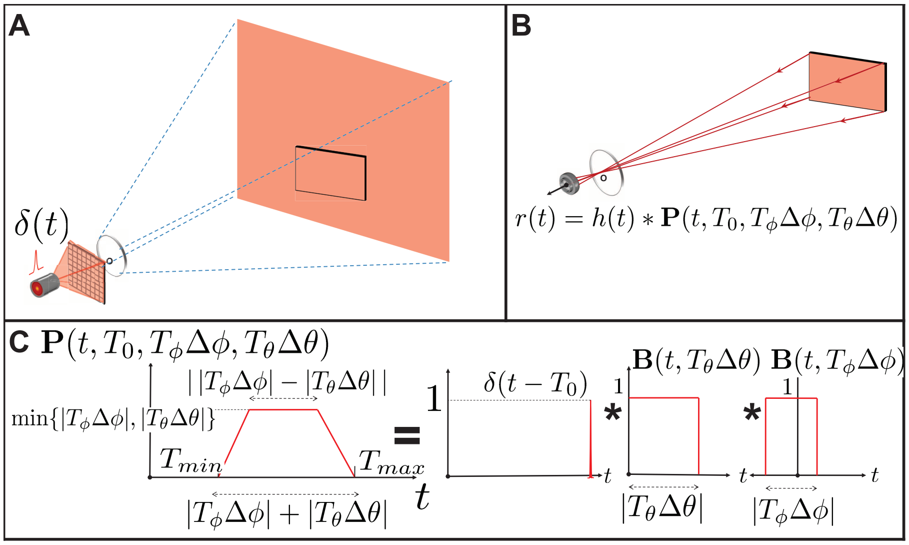
A
$\delta(t)$
B
$r(t) = h(t) * \mathbf{P}(t, T_0, T_\phi\Delta\phi, T_\theta\Delta\theta)$
C $\mathbf{P}(t, T_0, T_\phi\Delta\phi, T_\theta\Delta\theta)$
$||T_\phi\Delta\phi| - |T_\theta\Delta\theta||$
$\min\{|T_\phi\Delta\phi|, |T_\theta\Delta\theta|\}$
T_{min}
T_{max}
t
$|T_\phi\Delta\phi| + |T_\theta\Delta\theta|$
=
1
$\delta(t - T_0)$
t
*
$\mathbf{B}(t, T_\theta\Delta\theta)$
1
t
$|T_\theta\Delta\theta|$
*
$\mathbf{B}(t, T_\phi\Delta\phi)$
1
t
t
$|T_\phi\Delta\phi|$

Fig. 4. (A) All-ones scene illumination. (B) Scene response to all-ones scene illumination. (C) Diagrammatic explanation of the modeling of the parametric signal $p(t)$.

Next, we use the expression for $r_q(t)$ to model the response of the scene in illumination to a fully transparent SLM pattern (see Fig. 4). The signal $r(t)$ obtained in this case is the total light incident at the photodetector from all possible positions of Q on the rectangular facet:

$$r(t) \;=\; a\int_{\phi_1}^{\phi_2}\int_{\theta_1}^{\theta_2} h(t - 2\,|OQ(\phi,\theta)|)\,d\theta\,d\phi, \tag{1}$$

presuming a linear detector response. From Fig. 3 we note that $|OQ(\phi,\theta)| = d_\perp\sqrt{\sec^2\phi + \tan^2\theta}$. Thus, substituting in Eq. (1) we have

$$\begin{aligned} r(t) \;&=\; a\int_{\phi_1}^{\phi_2}\int_{\theta_1}^{\theta_2} h\left(t - 2d_\perp\sqrt{\sec^2\phi + \tan^2\theta}\right)\,d\theta\,d\phi \\ &=\; a\int_{0}^{\Delta\phi}\int_{0}^{\Delta\theta} h\left(t - 2d_\perp\sqrt{\sec^2(\phi_1+\phi) + \tan^2(\theta_1+\theta)}\right)\,d\theta\,d\phi, \end{aligned} \tag{2}$$

where the equality in Eq. (2) follows from a change of variables $\phi \leftarrow (\phi - \phi_1)$ and $\theta \leftarrow (\theta - \theta_1)$. Since $\theta \in [0, \Delta\theta]$ and $\phi \in [0, \Delta\phi]$ are small angles, $\sqrt{\sec^2(\phi_1+\phi) + \tan^2(\theta_1+\theta)}$ is approximated well using a first-order expansion:

$$\begin{aligned} &\sqrt{\sec^2(\phi_1+\phi) + \tan^2(\theta_1+\theta)} \\ &\approx\; \sqrt{\sec^2\phi_1 + \tan^2\theta_1} + \frac{1}{\sqrt{\sec^2\phi_1 + \tan^2\theta_1}}\left((\tan\phi_1\sec^2\phi_1)\,\phi + (\tan\theta_1\sec^2\theta_1)\,\theta\right). \end{aligned} \tag{3}$$

For notational simplicity, let $\gamma(\phi_1,\theta_1) = \sqrt{\sec^2\phi_1 + \tan^2\theta_1}$. Using Eq. (3), Eq. (2) is approximated well by

$$\begin{aligned} r(t) \;&=\; a\int_{0}^{\Delta\phi}\int_{0}^{\Delta\theta} h\left(t - 2d_\perp\left(\gamma(\phi_1,\theta_1) + \frac{(\tan\phi_1\sec^2\phi_1)\,\phi + (\tan\theta_1\sec^2\theta_1)\,\theta}{\gamma(\phi_1,\theta_1)}\right)\right)\,d\theta\,d\phi \\ &=\; a\int_{0}^{\Delta\phi}\int_{0}^{\Delta\theta} h(t - \tau(\phi,\theta))\,d\theta\,d\phi, \end{aligned}$$

Received 19 Aug 2011; revised 2 Oct 2011; accepted 3 Oct 2011; published 17 Oct 2011

where

$$\tau(\phi,\theta) \;=\; 2d_{\perp}\gamma(\phi_1,\theta_1) + \frac{2d_{\perp}}{\gamma(\phi_1,\theta_1)}(\tan\phi_1 \sec^2\phi_1)\phi + \frac{2d_{\perp}}{\gamma(\phi_1,\theta_1)}(\tan\theta_1 \sec^2\theta_1)\theta. \qquad (4)$$

We now make an important observation. The time delay function $\tau(\phi,\theta)$ is a linear function of the angular variations $\phi_1 \le \phi \le \phi_2$ and $\theta_1 \le \theta \le \theta_2$. Thus, the time-difference-of-arrival of the returns from the closest point of the rectangular facet to the farthest point varies linearly. This is the central observation that allows us to model the returned signal using a parametric signal processing framework (as discussed next) and recover the scene depth variations using the proposed acquisition setup. Again for notational simplicity, let

$$T_0 = 2d_{\perp}\gamma(\phi_1,\theta_1), \quad T_\phi = \frac{2d_{\perp}}{\gamma(\phi_1,\theta_1)}\tan\phi_1\sec^2\phi_1, \quad T_\theta = \frac{2d_{\perp}}{\gamma(\phi_1,\theta_1)}\tan\theta_1\sec^2\theta_1.$$

Note that $T_0 > 0$ for all values of ϕ_1 and θ_1, but T_ϕ and T_θ may be negative or positive. With this notation and a change of variables, $\tau_1 \leftarrow T_\phi\,\phi$ and $\tau_2 \leftarrow T_\theta\,\theta$, we obtain

$$\begin{aligned} r(t) &= a\int_0^{\Delta\phi}\int_0^{\Delta\theta} h\left(t - T_0 - T_\phi\,\phi - T_\theta\,\theta\right)d\theta\,d\phi \\ &= \frac{a}{T_\phi\,T_\theta}\int_0^{T_\phi\,\Delta\phi}\int_0^{T_\theta\,\Delta\theta} h\left(t - T_0 - \tau_1 - \tau_2\right)d\tau_1\,d\tau_2 \\ &= \frac{a}{T_\phi\,T_\theta}h(t) * \delta(t - T_0) * \int_0^{T_\phi\,\Delta\phi}\delta(t-\tau_1)\,d\tau_1 * \int_0^{T_\theta\,\Delta\theta}\delta(t-\tau_2)\,d\tau_2 \\ &= \frac{a}{T_\phi T_\theta}h(t) * \delta(t - T_0) * \mathbf{B}(t, T_\phi\,\Delta\phi) * \mathbf{B}(t, T_\theta\,\Delta\theta) \end{aligned}$$

where $\mathbf{B}(t,T)$ is the *box function* with width $|T|$ as shown in Fig. 4C and defined as

$$\mathbf{B}(t,T) \;=\; \begin{cases} 1, & \text{for } t \text{ between 0 and } T; \\ 0, & \text{otherwise.} \end{cases}$$

The function $\mathbf{B}(t,T)$ is a *parametric function* that can be described with a small number of parameters despite its infinite Fourier bandwidth [17, 18]. The convolution of $\mathbf{B}(t,T_\phi\,\Delta\phi)$ and $\mathbf{B}(t,T_\theta\,\Delta\theta)$, delayed in time by T_0, is another parametric function as shown in Fig. 4C. We call this function $\mathbf{P}(t,T_0,T_\phi\,\Delta\phi,T_\theta\,\Delta\theta)$. It is piecewise linear and plays a central role in our depth acquisition approach for piecewise-planar scenes. With this notation, we obtain

$$r(t) \;=\; \frac{a}{T_\phi T_\theta}h(t) * \mathbf{P}(t,T_0,T_\phi\,\Delta\phi,T_\theta\,\Delta\theta).$$

The function $\mathbf{P}(t,T_0,T_\phi\,\Delta\phi,T_\theta\,\Delta\theta)$ is nonzero over a time interval $t \in [T_{\min}, T_{\max}]$ that is precisely the time interval in which reflected light from the points on the rectangular planar facet arrives at the detector. Also, for intuition, note that T_0 is equal to the distance between O and the lower left corner of the rectangular plane, but it may or may not be the point on the plane closest to O. With knowledge of $T_{\min}$ and $T_{\max}$ we obtain a region of certainty in which the rectangular facet lies. This region is a spherical shell centered at O with inner and outer radii equal to $T_{\min}$ and $T_{\max}$ respectively (see Fig. 5). Within this shell, the rectangular planar facet may have many possible orientations and positions.

#153103 - $15.00 USD Received 19 Aug 2011; revised 2 Oct 2011; accepted 3 Oct 2011; published 17 Oct 2011

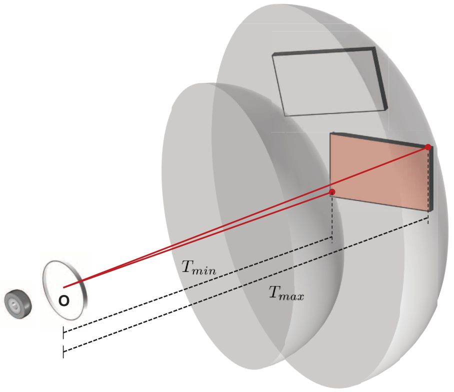

Fig. 5. The signal $p(t)$ only provides information regarding the depth ranges present in the scene. It does not allow us to estimate the position and shape of the planar facet in the FOV of the imaging system. At best, the facet can be localized to lie between spherical shells specified by $T_{\min}$ and $T_{\max}$. In this figure two possible positions for the rectangular facet are shown.

3.2. Parameter recovery

We wish to estimate the function $\mathbf{P}(t, T_0, T_\phi\,\Delta\phi, T_\theta\,\Delta\theta)$ and hence the values of $T_{\min}$ and $T_{\max}$ by processing the digital samples $r[k]$ of the function $r(t)$. The detector impulse response $h(t)$ is generally modeled as a bandlimited lowpass filter. Thus, the general deconvolution problem of obtaining $\mathbf{P}(t, T_0, T_\phi\,\Delta\phi, T_\theta\,\Delta\theta)$ from samples $r[k]$ is ill-posed and highly sensitive to noise. However, our modeling shows that the light transport function $\mathbf{P}(t, T_0, T_\phi\,\Delta\phi, T_\theta\,\Delta\theta)$ is piecewise linear. This knowledge makes the recovery of $\mathbf{P}(t, T_0, T_\phi\,\Delta\phi, T_\theta\,\Delta\theta)$ a *parametric deconvolution* problem that we solve using the parametric signal processing framework described in [27].

It is important to emphasize that the analysis up to this point is independent of the tilt α and orientation of the rectangular plane with respect to the global coordinate system (X,Y,Z); i.e., the tilt α has not appeared in any mathematical expression. Thus the parametric function $\mathbf{P}(t, T_0, T_\phi\,\Delta\phi, T_\theta\,\Delta\theta)$ describing the light transport between the imaging device and the rectangular planar facet is independent of the orientation of the line OC. This is intuitive because all the results were derived by considering a new frame of reference involving the rectangular plane and the normal to the plane from the origin, OC. The derived parametric light signal expressions themselves did not depend on how OC is oriented with respect to the global coordinate system but rather depend on the relative position of the plane with respect to OC. This explains why it is not possible to infer the position and orientation of the planar facet in the FOV of the system from the estimates of $\mathbf{P}(t, T_0, T_\phi\,\Delta\phi, T_\theta\,\Delta\theta)$. Recovery of the position and orientation of a rectangular planar facet is accomplished in Step 2 of our method using patterned illuminations as described in Section 4 below.

4. Response of a single rectangular facet to binary SLM pattern

4.1. Notation

As discussed in Section 2, the SLM pixels discretize the FOV into small squares of size $\Delta \times \Delta$. We index both the SLM pixels and the corresponding scene points by (i, j). Since we illuminate the scene with a series of M different binary SLM patterns, we also assign an index p for the

#153103 - $15.00 USD Received 19 Aug 2011; revised 2 Oct 2011; accepted 3 Oct 2011; published 17 Oct 2011

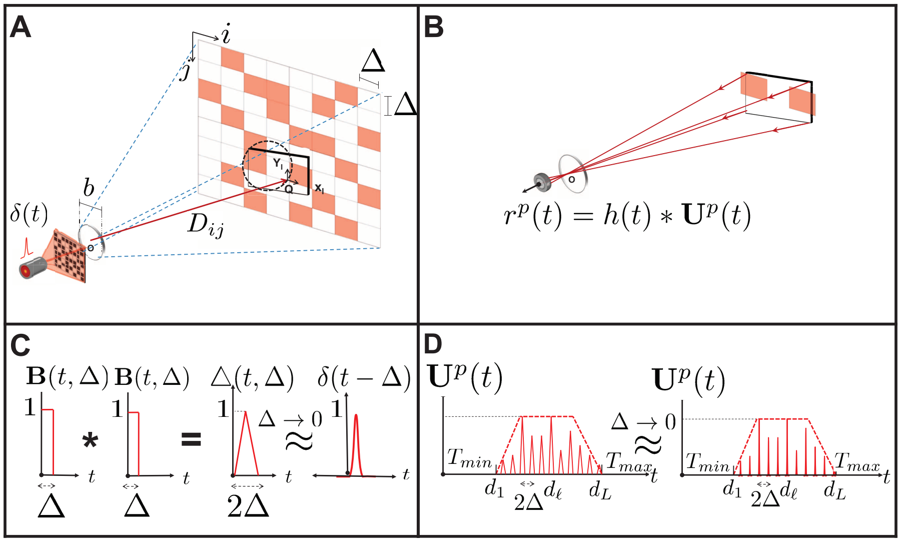

Fig. 6. (A) Binary patterned scene illumination. (B) Scene response to binary patterned scene illumination. (C) Diagrammatic explanation of the high-resolution SLM (small Δ) approximation. (D) Modeling of the parametric signal $\mathbf{U}^p(t)$ as a weighted sum of equally-spaced Diracs. Note that $\mathbf{U}^p(t)$ has the same time envelope as the signal $\mathbf{P}(t, T_0, T_\phi\,\Delta\phi, T_\theta\,\Delta\theta)$.

illumination patterns. The full collection of binary SLM values is denoted $\{c_{ij}^p : i = 1,\ldots,N,\ j = 1,\ldots,N,\ p = 1,\ldots,M\}$.

Let $\mathbf{D}$ denote the *depth map* that we wish to construct. Then $\mathbf{D}_{ij}$ is the depth in the direction of illumination of SLM pixel (i, j), assuming rays in that direction intersect the rectangular facet; set $\mathbf{D}_{ij}$ to zero otherwise. More specifically, we use the lower-left corner of the projection of the pixel onto the planar facet, as shown in Fig. 6A. It is convenient to also define the *index map*, $\mathbf{I} = \{\mathbf{I}_{ij} : i = 1,\ldots,N,\ j = 1,\ldots,N\}$, associated with the rectangular facet through

$$\mathbf{I}_{ij} = \begin{cases} 1, & \text{if rays along SLM illumination pixel } (i, j) \text{ intersect the rectangular facet;} \\ 0, & \text{otherwise.} \end{cases}$$

4.2. Scene response

If we consider the rectangular facet as being composed of smaller rectangular facets of size $\Delta \times \Delta$, then following the derivation described in Section 3.1 we find that the light signal received at the detector in response to patterned, impulsive illumination of the rectangular facet is given by

$$\begin{aligned} r^p(t) &= \sum_{i=1}^{N}\sum_{j=1}^{N} c_{ij}^p \mathbf{I}_{ij} \left(a\,h(t) * \int_0^\Delta \int_0^\Delta \delta(t - 2\mathbf{D}_{ij} - 2x_\ell - 2y_\ell)\,dx_\ell\,dy_\ell \right) \qquad (5) \\ &= \sum_{i=1}^{N}\sum_{j=1}^{N} c_{ij}^p \mathbf{I}_{ij} \left(\frac{a}{4} h(t) * \delta(t - 2\mathbf{D}_{ij}) * \mathbf{B}(t,\Delta) * \mathbf{B}(t,\Delta) \right) \\ &= \frac{a}{4} h(t) * \left(\sum_{i=1}^{N}\sum_{j=1}^{N} c_{ij}^p \mathbf{I}_{ij} \left(\delta(t - 2\mathbf{D}_{ij}) * \mathbf{B}(t,\Delta) * \mathbf{B}(t,\Delta) \right) \right). \qquad (6) \end{aligned}$$

Next, define the signal $\mathbf{U}^p(t)$ as

$$\mathbf{U}^p(t) \;=\; \sum_{i=1}^{N}\sum_{j=1}^{N} c_{ij}^{p}\mathbf{I}_{ij}\left(\delta(t-2\mathbf{D}_{ij}) * \mathbf{B}(t,\Delta) * \mathbf{B}(t,\Delta)\right). \tag{7}$$

The function $\triangle(t,\Delta) = \mathbf{B}(t,\Delta) * \mathbf{B}(t,\Delta)$ has a triangular shape with a base width of 2Δ as shown in Fig. 6C. In practice, when the SLM has high spatial resolution then Δ is very small, i.e., $\Delta \ll W$, $\Delta \ll L$, and $\triangle(t,\Delta)$ approximates a Dirac delta function $\delta(t)$. Thus, for a high-resolution SLM the signal $\mathbf{U}^p(t)$ is a weighted sum of uniformly-spaced impulses where the spacing between impulses is equal to 2Δ. Mathematically, we use $\lim_{\Delta\to 0}\mathbf{B}(t,\Delta) * \mathbf{B}(t,\Delta) \;=\; \lim_{\Delta\to 0}\delta(t-\Delta) \;=\; \delta(t)$ in Eq. (7) to obtain

$$\lim_{\Delta\to 0}\mathbf{U}^p(t) \;=\; \sum_{i=1}^{N}\sum_{j=1}^{N} c_{ij}^{p}\mathbf{I}_{ij}\left(\delta(t-2\mathbf{D}_{ij}) * \delta(t)\right) \;=\; \sum_{i=1}^{N}\sum_{j=1}^{N} c_{ij}^{p}\mathbf{I}_{ij}\,\delta(t-2\mathbf{D}_{ij}). \tag{8}$$

The parametric signal $\mathbf{U}^p(t)$ is obtained in the process of illuminating the scene with a patterned illumination and collecting light from illuminated portions of the scene ($c_{ij}^{p} = 1$) where the rectangular planar facet is present ($\mathbf{I}_{ij} = 1$). In particular, for a small value of Δ and fully-transparent SLM pattern (all-ones or $c_{ij}^{p} = 1 : i = 1,\ldots,N,\ j = 1,\ldots,N$) we have the following relation:

$$\begin{aligned} r^{\text{all-ones}}(t) &= \lim_{\Delta\to 0}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbf{I}_{ij}\left(a\,h(t) * \int_0^{\Delta}\int_0^{\Delta}\delta(t-2\mathbf{D}_{ij}-2x_\ell-2y_\ell)\,dx_\ell\,dy_\ell\right) && (9)\\ &= a\int_{\phi_1}^{\phi_2}\int_{\theta_1}^{\theta_2} h(t-2\,|OQ(\phi,\theta)|)\,d\theta\,d\phi = r(t) && (10)\end{aligned}$$

where Eq. (10) follows from the fact that the double-summation approximates the double integral in the limiting case ($\Delta \to 0$). Additionally, Eq. (10) implies that $\mathbf{U}^{\text{all-ones}}(t) = \mathbf{P}(t,T_0,T_\phi\,\Delta\phi,T_\theta\,\Delta\theta)$. An important observation that stems from this fact is that for any chosen illumination pattern, the signal $\mathbf{U}^p(t)$ and the signal $\mathbf{P}(t,T_0,T_\phi\,\Delta\phi,T_\theta\,\Delta\theta)$, which is obtained by using the all-ones or fully-transparent illumination pattern, have support in time $[T_{\min},\,T_{\max}]$. To be precise, if the points on the rectangular planar facet that are closest and farthest to O are illuminated, then both $\mathbf{U}^p(t)$ and $\mathbf{P}(t,T_0,T_\phi\,\Delta\phi,T_\theta\,\Delta\theta)$ have exactly the same duration and time delay. In practice, the binary patterns are randomly chosen with at least half of the SLM pixels "on," so it is highly likely that at least one point near the point closest to O and at least one point near the point farthest from O are illuminated. Hence, $\mathbf{U}^p(t)$ and $\mathbf{P}(t,T_0,T_\phi\,\Delta\phi,T_\theta\,\Delta\theta)$ are likely to have approximately the same time support and time delay offset. This implies $\mathbf{D}_{ij} \in [T_{\min},\,T_{\max}]$ (because the speed of light is normalized to unity).

4.3. Sampled data and Fourier-domain representation

Digital samples of the received signal $r^p[k]$ allow us to recover the depth map $\mathbf{D}$. First, note that the set of distance values, $\{\mathbf{D}_{ij} : i = 1,\ldots,N,\ j = 1,\ldots,N\}$, may contain repetitions; i.e., several (i,j) positions may have the same depth value $\mathbf{D}_{ij}$. All these points will lie on a circular arc on the rectangular facet as shown in Fig. 6A. Each $\mathbf{D}_{ij}$ belongs to the set of equally-spaced distinct depth values $\{d_1,d_2,\ldots,d_L\}$ where

$$L = \frac{T_{\max}-T_{\min}}{2\Delta}, \quad d_1 = T_{\min}, \quad d_\ell = d_1 + 2\Delta\ell, \quad \ell = 1,\ldots,L.$$

Note that the linear variation of the depths $d_1,\ldots,d_L$ is a direct consequence of Eq. (4), which states that there is a linear variation of distance from O of the closest point on the rectangular

#153103 - $15.00 USD Received 19 Aug 2011; revised 2 Oct 2011; accepted 3 Oct 2011; published 17 Oct 2011

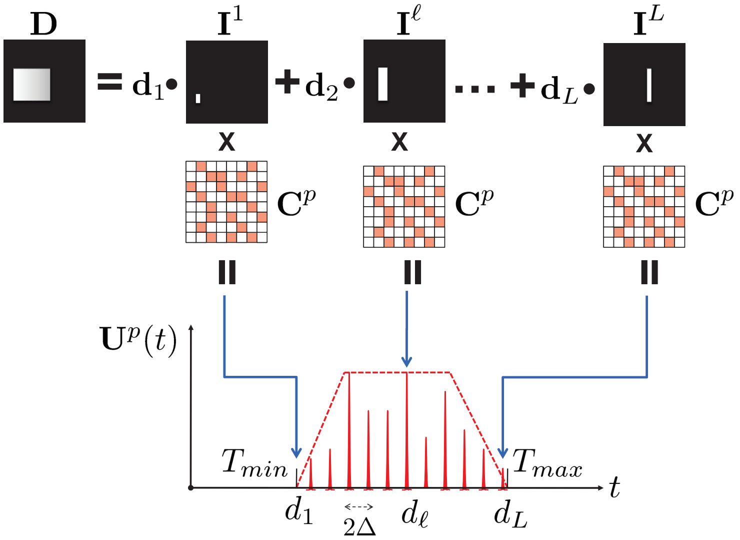

Fig. 7. Depth masks are binary-valued $N \times N$ pixel resolution images which indicate the presence (1) or absence (0) of a particular depth at a particular position (i, j) in the discretized FOV of the sensor. Depending on Δ and the sampling rate, we obtain a uniform sampling of the depth range and hence obtain L depth masks, one per depth value. The depth map of a scene is the weighted linear combination of depth masks where the weights are the numerical values of the discretized depth range, $\{d_1, d_2, \ldots, d_L\}$.

facet to the farthest. In the case of all-ones SLM illumination discussed in Section 3.1, we obtain the continuous signal $\mathbf{P}(t, T_0, T_\phi \Delta\phi, T_\theta \Delta\theta)$; in the patterned illumination case, we obtain a signal $\mathbf{U}^p(t)$ that is a weighted sum of uniformly-spaced impulses. With this new observation we have

$$\lim_{\Delta \to 0} \mathbf{U}^p(t) \;=\; \sum_{i=1}^{N}\sum_{j=1}^{N} c_{ij}^p \mathbf{I}_{ij}\, \delta(t - 2\mathbf{D}_{ij}) \;=\; \sum_{\ell=1}^{L} \left(\sum_{i=1}^{N}\sum_{j=1}^{N} c_{ij}^p I_{ij}^\ell \right) \delta(t - 2d_\ell), \tag{11}$$

where we define the matrix I^ℓ as

$$I_{ij}^\ell = \begin{cases} 1, & \text{if } \mathbf{D}_{ij} = d_\ell; \\ 0, & \text{otherwise,} \end{cases}$$

so $\mathbf{I}_{ij} = \sum_{\ell=1}^{L} I_{ij}^\ell$. and $\mathbf{D}_{ij} = \sum_{\ell=1}^{L} d_\ell I_{ij}^\ell$. With this new notation, the depth map $\mathbf{D}$ associated with the rectangular facet is the weighted sum of the index maps $\{I^\ell : \ell = 1, \ldots, L\}$ (see Fig. 7). Thus, constructing the depth map is now solved by finding the the L binary-valued index maps.

Taking the Fourier transform $\mathfrak{F}\{\cdot\}$ of the signals on both sides of Eq. (11) we get

$$\begin{aligned} \mathfrak{F}\left\{ \lim_{\Delta \to 0} \mathbf{U}^p(t) \right\} &= \mathfrak{F}\left\{ \sum_{\ell=1}^{L} \left(\sum_{i=1}^{N}\sum_{j=1}^{N} c_{ij}^p I_{ij}^\ell \right) \delta(t - 2d_\ell) \right\} \\ &= \sum_{\ell=1}^{L} \left(\sum_{i=1}^{N}\sum_{j=1}^{N} c_{ij}^p I_{ij}^\ell \right) \mathfrak{F}\{\delta(t - 2d_\ell)\} \;=\; \sum_{\ell=1}^{L} \left(\sum_{i=1}^{N}\sum_{j=1}^{N} c_{ij}^p I_{ij}^\ell \right) e^{-\mathbf{i}\omega 2 d_\ell} \end{aligned}$$

#153103 - $15.00 USD Received 19 Aug 2011; revised 2 Oct 2011; accepted 3 Oct 2011; published 17 Oct 2011

where $\mathbf{i} = \sqrt{-1}$. From elementary Fourier analysis and Eq. (6) we know that

$$\mathfrak{F}\{r^p(t)\} \;=\; \frac{a}{4}\mathfrak{F}\{h(t) * \mathbf{U}^p(t)\} \;=\; \frac{a}{4}\mathfrak{F}\{h(t)\}\,\mathfrak{F}\{\mathbf{U}^p(t)\}\,.$$

Let the ADC sample the signal incident on the photodetector at a sampling frequency of f samples per second. Then, using elementary sampling theory [28], we obtain the relation

$$\mathfrak{F}\{r^p[k]\} \;=\; \frac{af}{4}\mathfrak{F}\{h[k]\}\,\mathfrak{F}\{\mathbf{U}^p[k]\} \implies \frac{\mathfrak{F}\{r^p[k]\}}{\mathfrak{F}\{h[k]\}} \;=\; \frac{af}{4}\sum_{\ell=1}^{L}\left(\sum_{i=1}^{N}\sum_{j=1}^{N} c_{ij}^{p} I_{ij}^{\ell}\right) e^{-\mathbf{i}(4\pi f d_\ell)k}.$$

Let K denote the total number of samples collected by the ADC and let the discrete Fourier transform (DFT) of the samples $\{r^p[k] : k = 1,\ldots,K\}$ be denoted by $\{R^p[k] : k = 1,\ldots,K\}$. Similarly define $\{H^p[k] : k = 1,\ldots,K\}$ for the impulse response samples $\{h^p[k] : k = 1,\ldots,K\}$. Then

$$\frac{R^p[k]}{H[k]} \;=\; \frac{af}{4}\sum_{\ell=1}^{L}\left(\sum_{i=1}^{N}\sum_{j=1}^{N} c_{ij}^{p} I_{ij}^{\ell}\right) e^{-\mathbf{i}(4\pi f d_\ell)k}, \qquad k = 1,\ldots,K. \tag{12}$$

For notational simplicity let

$$y_\ell^p \;=\; \sum_{i=1}^{N}\sum_{j=1}^{N} c_{ij}^{p} I_{ij}^{\ell}, \qquad \ell = 1,\ldots,L. \tag{13}$$

The constants a and f are computed using calibration and are computationally compensated using normalization. Since the values $\{d_1, d_2, \ldots, d_L\}$ are known, Eq. (12) can be represented as a system of linear equations as follows:

$$\begin{bmatrix} R^p[1]/H[1] \\ \vdots \\ R^p[k]/H[k] \\ \vdots \\ R^p[K]/H[K] \end{bmatrix} = \begin{bmatrix} 1 & \cdots & 1 & \cdots & 1 \\ \vdots & & \vdots & & \vdots \\ e^{-\mathbf{i}(4\pi f d_1)k} & \cdots & e^{-\mathbf{i}(4\pi f d_\ell)k} & \cdots & e^{-\mathbf{i}(4\pi f d_L)k} \\ \vdots & & \vdots & & \vdots \\ e^{-\mathbf{i}(4\pi f d_1)K} & \cdots & e^{-\mathbf{i}(4\pi f d_\ell)K} & \cdots & e^{-\mathbf{i}(4\pi f d_L)K} \end{bmatrix} \begin{bmatrix} y_1^p \\ \vdots \\ y_\ell^p \\ \vdots \\ y_L^p \end{bmatrix},$$

which can be compactly written as

$$\mathbf{R}^p/\mathbf{H} = \mathbf{V}\mathbf{y}^p \tag{14}$$

(where the division is elementwise). The matrix V is a *Vandermonde* matrix; thus $K \geq L$ ensures that we can uniquely solve the linear system in Eq. (14). Furthermore, a larger value of K allows us to mitigate the effect of noise by producing least square estimates of $\mathbf{y}^p$.

Next, from Eq. (13) we see that $\mathbf{y}^p$ can also be represented with a linear system of equations as follows:

$$\begin{bmatrix} y_1^p \\ \vdots \\ y_\ell^p \\ \vdots \\ y_L^p \end{bmatrix} = \begin{bmatrix} I_{11}^{1} & \cdots & I_{1N}^{1} & I_{21}^{1} & \cdots & I_{2N}^{1} & \cdots & I_{N1}^{1} & \cdots & I_{NN}^{1} \\ \vdots & & \vdots & \vdots & & \vdots & & \vdots & & \\ I_{11}^{\ell} & \cdots & I_{1N}^{\ell} & I_{21}^{\ell} & \cdots & I_{2N}^{\ell} & \cdots & I_{N1}^{\ell} & \cdots & I_{NN}^{\ell} \\ \vdots & & \vdots & \vdots & & \vdots & & \vdots & & \\ I_{11}^{L} & \cdots & I_{1N}^{L} & I_{21}^{L} & \cdots & I_{2N}^{L} & \cdots & I_{N1}^{L} & \cdots & I_{NN}^{L} \end{bmatrix} \begin{bmatrix} c_{11}^p \\ \vdots \\ c_{1N}^p \\ c_{21}^p \\ \vdots \\ c_{2N}^p \\ \vdots \\ c_{N1}^p \\ \vdots \\ c_{NN}^p \end{bmatrix}. \tag{15}$$

From the M different binary SLM illumination patterns, we get M instances of Eq. (15) that can be combined into the compact representation

$$\underbrace{\mathbf{y}}_{L\times M} = \underbrace{\left[I^1 \cdots I^\ell \cdots I^L\right]^T}_{L\times N^2} \underbrace{\mathbf{C}}_{N^2\times M}. \tag{16}$$

This system of equations is under-constrained since there are $L \times N^2$ unknowns (corresponding to the unknown values of $\left[I^1 \ldots I^\ell \ldots I^L\right]$) and only $L \times M$ available transformed data observations $\mathbf{y}$. Note that $\mathbf{y}$ is computed using a total of $K \times M$ samples of the light signals received in response to $M \ll N^2$ patterned illuminations.

4.4. Algorithms for depth map reconstruction

Our goal is now to recover the depth map $\mathbf{D}$, which has $N \times N$ entries. To enable depth map reconstruction even though we have much fewer observations than unknowns, we exploit the structure of scene depth. We know that the depth values $\mathbf{D}_{ij}$ correspond to the distances from O to points that are constrained to lie on a rectangular facet and that the distances $\mathbf{D}_{ij}$ are also linearly spaced between d_1 and d_L. The planar constraint and linear variation imply that the depth map $\mathbf{D}$ is *sparse* in the second-finite difference domain as shown Fig. 2. By exploiting this sparsity of the depth map, it is possible to recover $\mathbf{D}$ from the data $\mathbf{y}$ by solving the following constrained ℓ_1-regularized optimization problem:

$$\begin{array}{lll}\textbf{OPT:} & \underset{\mathbf{D}}{\text{minimize}} & \left\|\mathbf{y} - \left[I^1 \ldots I^\ell \ldots I^L\right]^T \mathbf{C}\right\|_F^2 + \left\|\left(\mathbf{\Phi} \otimes \mathbf{\Phi}^T\right)\mathbf{D}\right\|_1 \\ & \text{subject to} & \displaystyle\sum_{\ell=1}^{L} I_{ij}^\ell = 1, \quad \text{for all } (i,j), \qquad \sum_{\ell=1}^{L} d_\ell I^\ell = \mathbf{D}, \qquad \text{and} \\ & & I_{ij}^\ell \in \{0,1\}, \quad \ell = 1,\ldots,L, \quad i = 1,\ldots,N, \quad j = 1,\ldots,N.\end{array}$$

Here the Frobenius matrix norm squared $\|.\|_F^2$ is the sum-of-squares of the matrix entries, the matrix $\mathbf{\Phi}$ is the second-order finite difference operator matrix

$$\mathbf{\Phi} = \begin{bmatrix} 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & -2 & 1 \end{bmatrix},$$

and $\otimes$ is the standard Kronecker product for matrices.

The optimization problem **OPT** has an intuitive interpretation. Our objective is to find the depth map $\mathbf{D}$ that is most consistent with having a piecewise-planar scene. Such scenes are characterized by $\mathbf{D}$ having a discrete two-dimensional Laplacian $\left(\mathbf{\Phi} \otimes \mathbf{\Phi}^T\right)\mathbf{D}$ with a small number of nonzero entries (corresponding to the boundaries of the planar facets). The number of nonzero entries (the "ℓ_0 pseudonorm") is difficult to use because it is nonconvex and not robust to small perturbations, and the ℓ_1 norm is a suitable proxy with many optimality properties [25]. The problem **OPT** combines the above objective with maintaining fidelity with the measured data by keeping $\|\mathbf{y} - \left[I^1 \ldots I^\ell \ldots I^L\right]\mathbf{C}\|_F^2$ small. The constraints $I_{ij}^\ell \in \{0, 1\}$ and $\sum_{\ell=1}^{L} I_{ij}^\ell = 1$ for all (i, j) are a mathematical rephrasing of the fact that each point in the depth map has a single depth value so different depth values cannot be assigned to one position (i, j). The constraint $\sum_{\ell=1}^{L} d_\ell I^\ell = \mathbf{D}$ expresses how the depth map is constructed from the index maps.

While the optimization problem **OPT** already contains a convex relaxation in its use of $\|\mathbf{\Phi}\mathbf{D}\|_1$, it is nevertheless computationally intractable because of the integrality constraints

#153103 - $15.00 USD Received 19 Aug 2011; revised 2 Oct 2011; accepted 3 Oct 2011; published 17 Oct 2011

$I_{ij}^{\ell} \in \{0, 1\}$. Using a further relaxation of $I_{ij}^{\ell} \in [0, 1]$ yields the following tractable formulation.

$$\begin{aligned}
\textbf{R-OPT:} \quad & \underset{\mathbf{D}}{\text{minimize}} \quad \left\| \mathbf{y} - \left[I^1 \ldots I^{\ell} \ldots I^L \right]^T \mathbf{C} \right\|_{\mathrm{F}}^2 + \left\| \left(\Phi \otimes \Phi^T \right) \mathbf{D} \right\|_1 \\
& \text{subject to} \quad \sum_{\ell=1}^{L} I_{ij}^{\ell} = 1, \quad \text{for all } (i,j), \qquad \sum_{\ell=1}^{L} d_{\ell} I^{\ell} = \mathbf{D}, \qquad \text{and} \\
& \qquad\qquad\quad I_{ij}^{\ell} \in [0,1] \quad \ell = 1, \ldots, L, \quad i = 1, \ldots, N, \quad j = 1, \ldots, N.
\end{aligned}$$

We solved the convex optimization problem **R-OPT** using CVX, a package for specifying and solving convex programs [29, 30].

Summarizing, the procedure for reconstructing the depth map of a scene with a single rectangular planar facet is as follows:

1. Measure the digital samples of the impulse response of the photodetector $\{h[k] : k = 1, \ldots, K\}$. We assume that the ADC samples at least twice as fast as the bandwidth of the photodetector (Nyquist criterion).
2. Illuminate the entire scene with an impulse using an all-ones, fully-transparent SLM pattern and measure the digital samples of the received signal $\{r[k] : k = 1, \ldots, K\}$. In case the source is periodic, such as an impulse train, the received signal $r(t)$ will also be periodic and hence the samples need to be collected only in one period.
3. Process the received signal samples $\{r[k] : k = 1, \ldots, K\}$ and the impulse response samples, $\{h[k] : k = 1, \ldots, K\}$ using the parametric signal deconvolution algorithm described in [27] to estimate the piecewise-linear function $\mathbf{P}(t, T_0, T_{\phi}\,\Delta\phi, T_{\theta}\,\Delta\theta)$.
4. Using the estimate of $\mathbf{P}(t, T_0, T_{\phi}\,\Delta\phi, T_{\theta}\,\Delta\theta)$, infer the values of $T_{\min}$ and $T_{\max}$.
5. Illuminate the scene $M = N^2/20$ times using the randomly-chosen binary SLM patterns $\{c_{ij}^{p} : p = 1, \ldots, M\}$, again using an impulsive light source. Record K digital time samples of the light signal received at the photodetector in response to each of the patterned illuminations $\{r^p[k] : k = 1, \ldots, K,\ p = 1, \ldots M\}$.
6. For each pattern, compute the transformed data $\mathbf{y} = [\mathbf{y}^1, \ldots, \mathbf{y}^M]$ as described in Section 4.2.
7. Construct the matrix $\mathbf{C}$ from the binary SLM patterns.
8. Solve the problem **R-OPT** to reconstruct the depth map $\mathbf{D}$ associated with the rectangular facet. This depth map contains information about the position, orientation and shape of the planar facet.

5. Depth map acquisition for general scenes

In this section we generalize the received signal model and depth map reconstruction developed in Sections 3 and 4 to planar facets of any shape and scenes with multiple planar facets.

5.1. *General planar shapes*

The signal modeling described in Section 3.1 applies to a planar facet with non-rectangular shape as well. For example, consider the illumination of a single triangular facet with the fully transparent SLM pattern as shown in Fig. 8 (left panel). In this case, the light signal received at the detector is

$$r(t) \;=\; a \int_{\phi_1}^{\phi_2} \int_{\theta_1(\phi)}^{\theta_2(\phi)} h(t - 2\,|OQ(\phi,\theta)|)\, d\theta\, d\phi.$$

#153103 - $15.00 USD Received 19 Aug 2011; revised 2 Oct 2011; accepted 3 Oct 2011; published 17 Oct 2011

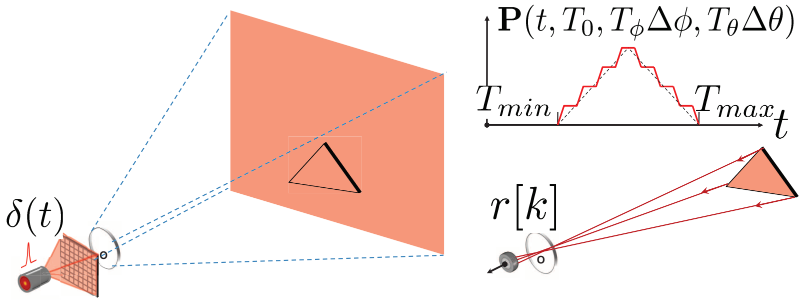
$\mathbf{P}(t, T_0, T_\phi\Delta\phi, T_\theta\Delta\theta)$
T_{min}
T_{max}
t
$\delta(t)$
$r[k]$

Fig. 8. Parametric modeling for non-rectangular planes. The piecewise linear fit (shown in dotted black) is a good fit to the true parametric scene response from a triangular planar facet. This fit allows us to robustly estimate $T_{\min}$ and $T_{\max}$.

Contrasting with Eq. (1), since the shape is not a rectangle, the angle θ does not vary over the entire range $[\theta_1, \theta_2]$. Instead, for a fixed value of angle ϕ, the angle θ can only vary from between some $\theta_1(\phi)$ and some $\theta_2(\phi)$. These limits of variation are determined by the shape of the object as shown in Fig. 8 (right panel).

Since the planar facet is in the far field, the distances of plane points from O still vary linearly. As a result, $r(t)$ is still equal to the convolution of the detector impulse response with a parametric signal whose shape depends on the shape of the planar facet. For example, as shown in Fig. 8 (right panel), the profile of the signal $\mathbf{P}(t, T_0, T_\phi\,\Delta\phi, T_\theta\,\Delta\theta)$ is triangular with jagged edges. The task of estimating the signal $\mathbf{P}(t, T_0, T_\phi\,\Delta\phi, T_\theta\,\Delta\theta)$ corresponding to a general shape, such as a triangle, from the samples $r[k]$ is more difficult than estimating $\mathbf{P}(t, T_0, T_\phi\,\Delta\phi, T_\theta\,\Delta\theta)$ in the case of a rectangular facet. However, as we can see from Fig. 8 (right panel), a good piecewise-linear fit is still obtained using the samples of $r[k]$. This piecewise-linear approximation, although not exact, suffices for our purpose of estimating the shortest and farthest distance to the points on the planar facet. Thus it is possible to estimate the values $T_{\min}$ and $T_{\max}$ using the samples $r[k]$ without any dependence on the shape of the planar facet. Once $T_{\min}$ and $T_{\max}$ are estimated, we use the framework described in Section 4 to recover the depth map of the scene, which will also reveal the exact shape and orientation of the planar facet.

5.2. *Multiple planar facets*

When the scene has multiple planar facets, as shown in Fig. 9-A, the linearity of light transport and the linear response of the detector together imply that the detector output is the sum of the signals received from each of the individual planar facets. This holds equally well for the cases of fully-transparent and patterned SLM illumination.

Figure 9A illustrates a scene composed of two planar facets illuminated with a fully-transparent SLM setting. The total response is given by

$$r(t) \;=\; r_1(t) + r_2(t) \;=\; \mathbf{P}_1(t, T_{0,1}, T_{\phi,1}\Delta\phi_1, T_{\theta,1}\Delta\theta_1) + \mathbf{P}_2(t, T_{0,2}, T_{\phi,2}\Delta\phi_2, T_{\theta,2}\Delta\theta_2),$$

where $r_i(t)$ and $\mathbf{P}_i$ denote the response from planar facet i. The total response is thus a parametric signal. When points on two different planar facets are at the same distance from O (see Fig. 9C), there is time overlap between $\mathbf{P}_A(t, T_{0_A}, T_{\phi_A}\Delta\phi_A, T_{\theta_A}\Delta\theta_A)$ and $\mathbf{P}_B(t, T_{0_B}, T_{\phi_B}\Delta\phi_B, T_{\theta_B}\Delta\theta_B)$ (see Fig. 9E). In any case, closest distance $T_{\min}$ and farthest distance $T_{\max}$ can be estimated from $r(t)$. Thus the framework developed in Section 4 for estimating the distance set $\{d_1, d_2, \ldots, d_L\}$ applies here as well. Note that we do not need any prior information

#153103 - $15.00 USD Received 19 Aug 2011; revised 2 Oct 2011; accepted 3 Oct 2011; published 17 Oct 2011

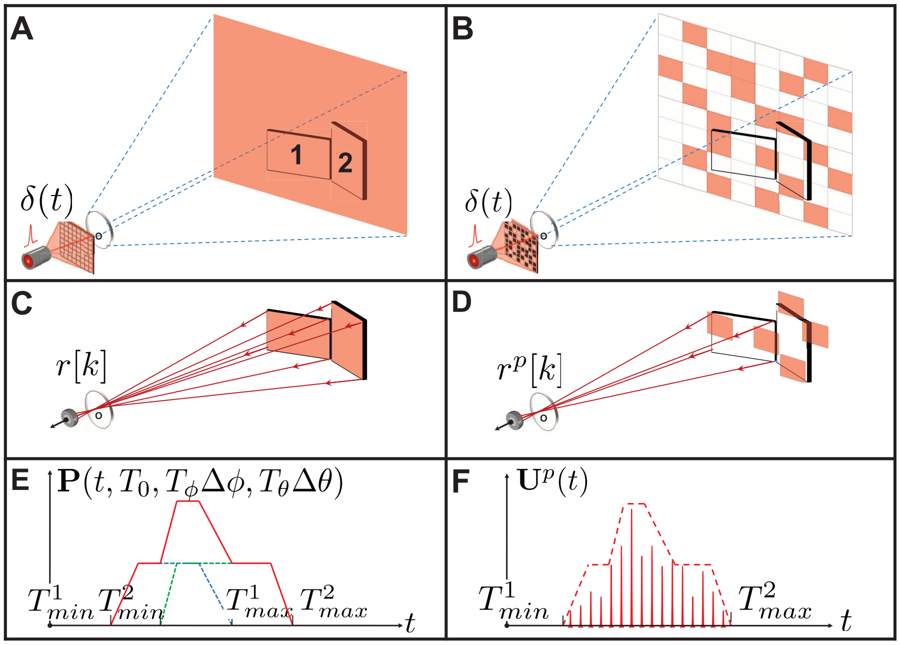
A
$\delta(t)$
1
2
B
$\delta(t)$
C
$r[k]$
D
$r^p[k]$
E
$\mathbf{P}(t, T_0, T_\phi\Delta\phi, T_\theta\Delta\theta)$
T^1_{min} T^2_{min} T^1_{max} T^2_{max} t
F
$\mathbf{U}^p(t)$
T^1_{min} T^2_{max} t

Fig. 9. Parametric modeling in scenes with multiple planar facets. Since light transport is linear and assuming light adds linearly at the detector, the parametric signal that characterizes the scene response is the sum of multiple parametric signals. Thus even in the case of multiple planar facets, a piecewise-linear fit to the observed data allows us to reliably estimate the scene's depth range.

on how many planar facets are present in the scene.

Figure 9B illustrates the same scene illuminated with a patterned SLM setting. Since the response to pattern p follows

$$r^p(t) \;=\; r^p_1(t) + r^p_2(t),$$

where $r^p_i(t)$ is the response from planar facet i, we can similarly write

$$\mathbf{U}^p(t) \;=\; \mathbf{U}^p_1(t) + \mathbf{U}^p_2(t).$$

Thus the problem of depth map reconstruction in case of scenes constituted of multiple planar facets is also solved using the convex optimization framework described in Section 4.

Figure 9 illustrates rectangular facets that do not occlude each other, but the lack of occlusion is not a fundamental limitation. If a portion of a facet is occluded, it effectively becomes nonrectangular, as described in Section 5.1.

6. Experiments

6.1. *Imaging setup and measurement*

The proof-of-concept experiment to demonstrate the single-sensor compressive depth acquisition framework is illustrated in Fig. 10. The periodic light source was a mode-locked Ti:Sapphire femtosecond laser with a pulse width of 100 fs and a repetition rate of 80 MHz operating at a wavelength of 790 nm. It illuminated a MATLAB-controlled Boulder Nonlinear

#153103 - $15.00 USD Received 19 Aug 2011; revised 2 Oct 2011; accepted 3 Oct 2011; published 17 Oct 2011

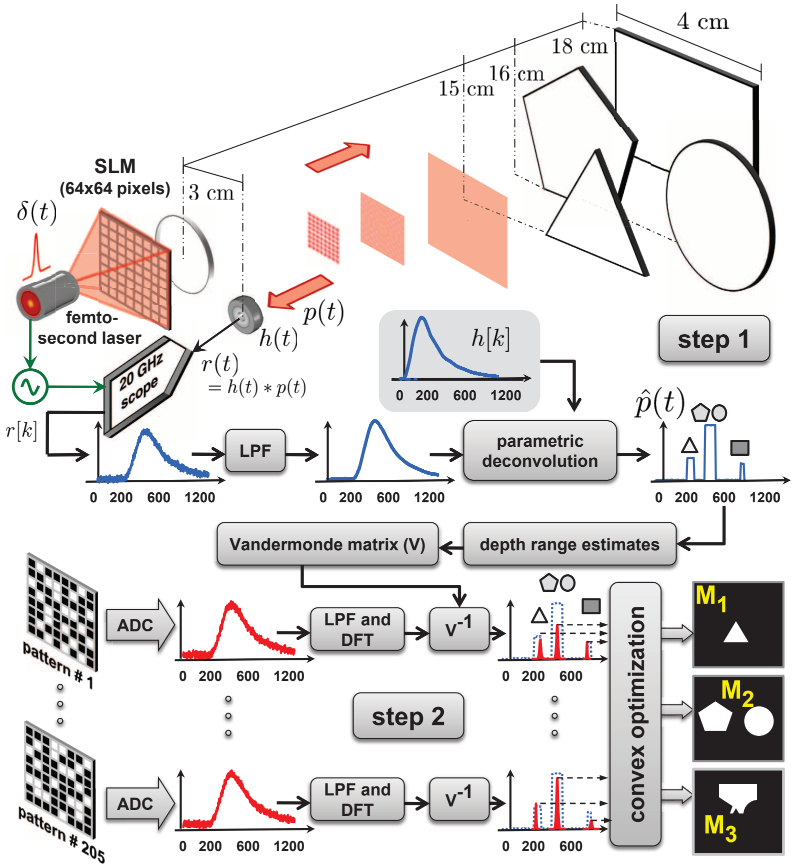

Fig. 10. Schematic experimental setup to demonstrate depth estimation using our proposed framework. See text for details.

Systems liquid-crystal SLM with a pixel resolution of 512×512 pixels, each 15×15 μm. Pixels were grouped in blocks of 8×8 and each block phase-modulated the incident light to either 0° or 180° phase. The phase-modulated beam was passed through a half-wave plate followed by a polarizer to obtain the binary intensity pattern. A total of 205 binary patterns of 64×64 block-pixel resolution, were used for illumination. Each pattern was randomly chosen and had about half of the 4096 SLM blocks corresponding to zero phase (zero intensity after the polarizer). The average power in an illumination pattern was about 40 to 50 mW. The binary patterns were serially projected onto the scene comprised of two to four Lambertian planar shapes (see Fig. 11A) at different inclinations and distances. Our piecewise-planar scenes were composed of acrylic cut-outs of various geometric shapes coated with Edmund Optics NT83-889 white reflectance coating. The effects of speckle and interference were minimized by using convex lenses to project the SLM patterns on the scene. At a distance of 10 cm from the detector, each pixel in the scene was about 0.1 mm^2. For each pattern, the light reflected from all the illuminated portions of the scene was focused on a ThorLabs DET10A Si PIN diode with a rise time of 0.7 ns and an active area of 0.8 mm^2. A transparent glass slide was used to direct a small portion of the light into a second photodetector to trigger a 20 GHz oscilloscope and obtain the time origin for all received signals.

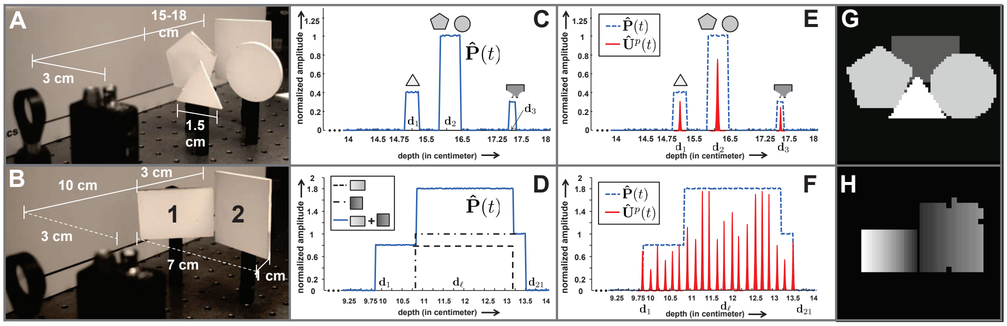

Fig. 11. Photographs of experimental setup (A and B). Parametric signal estimate in response to all-transparent illumination (C and D). Parametric signal estimate in response to patterned illumination (E and F). Depth map reconstructions (G and H).

The depth map recovery is a two-step process: first we estimate the depth range within which the scene is present, and then we estimate the spatial locations, orientations and shapes of the planar facets. In Step 1, the scene was first illuminated with an all-ones pattern. The resulting convolution, $r(t)$, of the scene's true parametric response $\mathbf{P}(t)$ and the detector's impulse response $h(t)$ was time sampled using the 20 GHz oscilloscope to obtain 1311 samples. These samples, $r[k]$, are lowpass filtered (LPF) to reduce sensor noise and processed using parametric deconvolution [17, 27, 31] to obtain the estimate $\hat{\mathbf{P}}(t)$ and hence the estimates of the distance ranges in which the planar facets lie. In Step 2, to recover the shapes and positions of the planar shapes, the scene is illuminated with 205 (5% of $64 \times 64 = 4096$) randomly-chosen binary patterns. The time samples collected in response to each patterned illumination are again low pass filtered (LPF) for denoising. The DFT of the filtered samples is processed using the Vandermonde matrix constructed using range estimates obtained in Step 1, to yield as many coefficients as there are distinct depth ranges (three in Fig. 10). These coefficients correspond to the product of the projected pattern and a binary-valued depth mask ($\mathbf{M_1}$, $\mathbf{M_2}$ and $\mathbf{M_3}$) that identifies the locations in the scene where the particular depth (d_1, d_2 and d_3) is present (see Fig. 7). The resulting 205×3 estimated coefficients are processed using a convex optimization framework that exploits the sparsity of the Laplacian of the depth map to recover the positions and shapes of the planar objects relative to the acquisition setup in the form of the three depth masks. Finally, these depth masks are weighted with the true depth values from Step 1 to reconstruct complete scene depth maps.

6.2. Depth map reconstruction results

Figures 11A and 11B show the relative positions and approximate distances between the SLM focusing lens, the photodetector, and the two scenes constituted of white colored, Lambertian planar facets of different shapes and sizes. In Fig. 11A (also see Fig. 10), the dimensions of the planar facets are about 10 times smaller than the separation between SLM/photodetector and scene. Thus, there is little variation in the times-of-arrival of reflections from points on any single planar facet, as evidenced by the three concentrated rectangular pulses in the estimated parametric signal $\hat{\mathbf{P}}(t)$ in Fig. 11C. The time delays correspond to the three distinct depth ranges (15 cm, 16 cm and 18 cm). In Fig. 11B, there is significant variation in the times-of-arrival of reflections from points within each planar facet as well as overlap in the returns from the two facets. Thus, we get a broader estimated parametric signal $\hat{\mathbf{P}}(t)$ that does not consist of disjoint rectangular pulses, and hence a continuous depth range as shown in Fig. 11D (solid blue curve). Overlaid on the experimental data in Fig. 11D are the computed separate contributions

#153103 - $15.00 USD Received 19 Aug 2011; revised 2 Oct 2011; accepted 3 Oct 2011; published 17 Oct 2011

from the two planes in Fig. 11B (black dashed and black dash-dotted curves), conforming to our modeling in Section 3. Note that the depth range axis is appropriately scaled to account for ADC sampling frequency and the factor of 2 introduced due to light going back and forth. The normalized amplitude of the parametric signal $\hat{\mathbf{P}}(t)$ is an approximate measure of how much surface area of the scene is at a particular depth. The depth discretization and hence the range resolution is governed by the size of the projected SLM pixel, Δ. In our experiment the measured Δ is 0.1 mm and hence there are 21 discrete depths, $d_1, \ldots, d_{21}$ at a separation of 2Δ. Fig. 11E and Fig. 11F show the parametric signal $\mathbf{U}^p(t)$ that is recovered in the case of the first patterned illumination for the scenes in Fig. 11A and Fig. 11B, respectively. Figs. 11G and 11H show 64×64-pixel depth maps reconstructed using time samples from patterned binary illuminations of both the scenes. The distinct depth values are rendered in gray scale with closest depth shown in white and farthest depth value shown in dark gray; black is used to denote the scene portions from where no light is collected.

Our technique yielded accurate sub-cm depth maps with sharp edges. The range resolution of our acquisition method—the ability to resolve close depths—depends on the bandwidth of the temporal light modulation, the response time of the photodetector, and the sampling rate of the ADC. The spatial resolution of our output depth map is a function of the number of distinct patterned scene illuminations; a complex scene with a large number of sharp features requires a larger number of SLM illuminations. In the presence of synchronization jitter and sensor noise, we average over multiple periods and use a larger number of illumination patterns to mitigate the effect of noise (see Fig. 10).

7. Discussion and extensions

The central novelty of our work relative to common LIDAR and TOF camera technologies is our mechanism for attaining spatial resolution through spatially-patterned illumination. In principle, this saves time relative to a LIDAR system because an SLM pattern can be changed more quickly than a laser position, and the number of acquisition cycles M is far fewer than the number of pixels in the constructed depth map. The savings relative to a TOF camera is in the number of sensors.

Our proposed depth acquisition technique also has two significant potential advantages over TOF cameras: First, our method is invariant to ambient light because only the low-frequency components of the recorded signals are affected by ambient light; low-frequency disturbances in turn only affect the overall scaling and do not affect the shape, duration and time delay of the parametric signal $\mathbf{P}(t)$. Second, there is potential for power savings: instead of constantly illuminating the scene with high-powered LED sources independent of the scene depth range, as is the case in TOF cameras, the scene range estimate from Step 1 of our method can be used to adaptively control the optical power output depending on how close the scene is to the imaging device.

The main limitation of our framework is inapplicability to scenes with curvilinear objects, which would require extensions of the current mathematical model. If we abandon the parametric signal recovery aspect of Step 1, we may still more crudely estimate the overall range of depths in the scene and proceed with Step 2. However, this will increase L and thus increase the computational complexity of depth map recovery. The degree to which it necessitates an increase in M requires further study. More generally, the relationship between M and the depth map quality requires further study; while the optimization problems introduced in Section 4.4 bear some similarity to standard compressed sensing problems, existing theory does not apply directly.

Another limitation is that a periodic light source creates a wrap-around error as it does in other TOF devices [7]. For scenes in which surfaces have high reflectance or texture variations,

#153103 - $15.00 USD Received 19 Aug 2011; revised 2 Oct 2011; accepted 3 Oct 2011; published 17 Oct 2011

availability of a traditional 2D image prior to our data acquisition allows for improved depth map reconstruction as discussed next.

7.1. Scenes with non-uniform texture and reflectance

Natural objects typically have surface texture and reflectance variations. In our experiments we only considered objects with uniform Lambertian reflectance. Here we briefly discuss the extension of our formulation to the case of planar facets with non-uniform texture and reflectance patterns. This extension assumes an SLM with a high number of pixels (small Δ) that performs grayscale light modulation. (Our experiments use only binary light modulation.)

Let the scene reflectance coefficient in the (i,j) direction be a_{ij}. Then the response to an all-ones (fully-transparent) SLM illumination is

$$\begin{aligned} r^0(t) &= \lim_{\Delta\to 0} \sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij} \mathbf{I}_{ij} \left(h(t) * \int_0^{\Delta} \int_0^{\Delta} \delta(t - 2\mathbf{D}_{ij} - 2x_l - 2y_l)\, dx_\ell \, dy_\ell \right) \\ &= \int_{\phi_1}^{\phi_2} \int_{\theta_1}^{\theta_2} a(\phi,\theta)\, h(t - 2\,|OQ(\phi,\theta)|)\, d\theta \, d\phi. \end{aligned}$$

The presence of the unknown reflectance variations $a(\phi,\theta)$ prevents us from modeling $r^0(t)$ as a convolution of $h(t)$ and a piecewise-linear parametric signal as described in Section 3.1. However, if prior to data acquisition we have a conventional 2D image (photograph) of the scene that provides an estimate of the scene reflectance $\{a_{ij} : i = 1,\ldots,N,\ j = 1,\ldots,N\}$, it is possible to compensate for the reflectance using a grayscale SLM illumination. Specifically, the "inverse" illumination pattern a/a_{ij}, $i = 1,\ldots,N$, $j = 1,\ldots,N$, where a is a chosen proportionality constant, yields response

$$\begin{aligned} r^{-1}(t) &= \lim_{\Delta\to 0} \sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij} \frac{a}{a_{ij}} \mathbf{I}_{ij} \left(h(t) * \int_0^{\Delta} \int_0^{\Delta} \delta(t - 2\mathbf{D}_{ij} - 2x_\ell - 2y_\ell)\, dx_\ell \, dy_\ell \right) \\ &= a \int_{\phi_1}^{\phi_2} \int_{\theta_1}^{\theta_2} h(t - 2\,|OQ(\phi,\theta)|)\, d\theta \, d\phi \;=\; h(t) * \mathbf{P}(t, T_0, T_\phi \,\Delta\phi, T_\theta \,\Delta\theta), \end{aligned}$$

suitable for Step 1 of our method. Analogous inversion of the scene reflectance can be applied in Step 2 of our method.

7.2. Use of non-impulsive illumination sources

In our formulation and experiments we used a light impulse generator such as a femtosecond laser as our illumination source. However, we note that since the photodetector impulse response $h(t)$ is bandlimited, the overall imaging system is bandlimited. Thus it is possible to use non-impulsive sources that match the band limit of the detector without losing any imaging quality. Here we derive an expression for the signal received at the photodetector when we use a general time-varying source $s(t)$ instead of an impulse $\delta(t)$.

The scene defines a linear and time-invariant (LTI) system from illumination to detection. This is easy to verify: light transport is linear, and if we illuminate the scene with a time-delayed pulse, the received signal is delayed by the same amount. We have already modeled as $r(t)$ the output of the system in response to impulse illumination. Thus, the signal received at the photodetector in response to illumination using source $s(t)$ is given by $s(t) * r(t)$, the convolution of $r(t)$ with the source signal $s(t)$. Since $r(t) = h(t) * \mathbf{P}(t, T_0, T_\phi \,\Delta\phi, T_\theta \,\Delta\theta)$ we have

$$s(t) * r(t) = s(t) * \{h(t) * \mathbf{P}(t, T_0, T_\phi \,\Delta\phi, T_\theta \,\Delta\theta)\} = \{s(t) * h(t)\} * \mathbf{P}(t, T_0, T_\phi \,\Delta\phi, T_\theta \,\Delta\theta). \quad (17)$$

Eq. (17) demonstrates that if we use a non-impulsive source $s(t)$ then all our formulations developed in Sections 3 and 4 are valid with one small change: use $s(t) * h(t)$ in place of $h(t)$.

8. Conclusion

We have presented a method for acquiring 2D depth maps of piecewise-planar scenes using time samples measured by a single photodetector in response to a series of spatiotemporally-modulated scene illuminations. In contrast to the moving 2D laser scanning in LIDAR systems and the focused 2D sensor array in TOF cameras, our acquisition architecture consists of a non-scanning 2D SLM and a single photodetector. We have demonstrated that it is possible to acquire scene depth at both high range resolution and spatial resolution with significantly-reduced device complexity and hardware cost as compared to state-of-the-art LIDAR systems and TOF cameras. We achieved these gains by developing a depth acquisition framework based on parametric signal modeling and sparsity of the Laplacian of the depth map of a typical scene.

Acknowledgments

This material is based upon work supported in part by the National Science Foundation under Grant No. 0643836 and by the DARPA InPho program through the US Army Research Office award W911-NF-10-1-0404. This work was also supported in part by NSF Major Research Instrumentation Program, Grant PHY-0959057, under the American Recovery and Reinvestment Act of 2009.

#153103 - $15.00 USD Received 19 Aug 2011; revised 2 Oct 2011; accepted 3 Oct 2011; published 17 Oct 2011