High-speed compressive range imaging based on active illumination

YANGYANG SUN,[1] XIN YUAN,[2] AND SHUO PANG[1,*]

[1]*College of Optics and Photonics (CREOL), University of Central Florida, 4304 Scorpius St. Orlando, FL 32816, USA*
[2]*Bell Labs, Nokia, 600 Mountain Avenue, Murray Hill, NJ 07974, USA*
[*]*pang@creol.ucf.edu*

Abstract: We report a compressive imaging method based on active illumination, which reconstructs a 3D scene at a frame rate beyond the acquisition speed limit of the camera. We have built an imaging prototype that projects temporally varying illumination pattern and demonstrated a joint reconstruction algorithm that iteratively retrieves both the range and high-temporal-frequency information from the 2D low-frame rate measurement. The reflectance and depth-map videos have been reconstructed at 1000 frames per second (fps) from the measurement captured at 200 fps. The range resolution is in agreement with the resolution calculated from the triangulation methods based on the same system geometry. We expect such an imaging method could become a simple solution to a wide range of applications, including industrial metrology, 3D printing, and vehicle navigations.

OCIS codes: (110.1758) Computational imaging; (150.5670) Range finding.

References and links

1. P. Treleaven and J. Wells, "3D body scanning and healthcare applications," Computer **40**(7), 28–34 (2007).
2. A. Corns and R. Shaw, "High resolution 3-dimensional documentation of archaeological monuments & landscapes using airborne lidar," J. Cult. Herit. **10**, e72 (2009).
3. F. Rengier, A. Mehndiratta, H. von Tengg-Kobligk, C. M. Zechmann, R. Unterhinninghofen, H. U. Kauczor, and F. L. Giesel, "3D printing based on imaging data: review of medical applications," Int. J. CARS **5**(4), 335–341 (2010).
4. W. S. Wijesoma, K. R. S. Kodagoda, and A. P. Balasuriya, "Road-boundary detection and tracking using ladarladar sensing," IEEE Trans. Robot. Autom. **20**(3), 456–464 (2004).
5. M. Hebert, "Active and passive range sensing for robotics," in Proc. 2000 ICRA. Millenn. Conf. IEEE Int. Conf. Robot. Autom. Symp. Proc. **1**, 102–110 (2000).
6. P. J. Besl, "Active, optical range imaging sensors," Mach. Vis. Appl. **1**(2), 127–152 (1988).
7. F. Blais, "Review of 20 years of range sensor development," J. Electron. Imaging **13**(1), 231 (2004).
8. P. Llull, X. Yuan, L. Carin, and D. J. Brady, "Image translation for single-shot focal tomography," Optica **2**(9), 822 (2015).
9. S. R. P. Pavani and R. Piestun, "High-efficiency rotating point spread functions," Opt. Express **16**(5), 3484–3489 (2008).
10. D. Gabor, "A new microscopic principle," Nature **161**(4098), 777–778 (1948).
11. D. L. Donoho, "Compressed sensing," IEEE Trans. Inf. Theory **52**(4), 1289–1306 (2006).
12. P. Llull, X. Liao, X. Yuan, J. Yang, D. Kittle, L. Carin, G. Sapiro, and D. J. Brady, "Coded aperture compressive temporal imaging," Opt. Express **21**(9), 10526–10545 (2013).
13. X. Yuan, P. Llull, X. Liao, J. Yang, D. J. Brady, and L. Carin, "Low-cost compressive sensing for color video and depth," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2014), pp. 3318–3325.
14. X. Yuan and S. Pang, "Structured illumination temporal compressive microscopy," Biomed. Opt. Express **7**(3), 746–758 (2016).
15. T.-H. Tsai, P. Llull, X. Yuan, L. Carin, and D. J. Brady, "Spectral-temporal compressive imaging," Opt. Lett. **40**(17), 4054–4057 (2015).
16. X. Yuan, "Generalized alternating projection based total variation minimization for compressive sensing," in Proceedings of the IEEE Conference on Image Processing (2016), pp. 2539.
17. J. Yang, X. Yuan, X. Liao, P. Llull, D. J. Brady, G. Sapiro, and L. Carin, "Video compressive sensing using Gaussian mixture models," IEEE Trans. Image Process. **23**(11), 4863–4878 (2014).
18. X. Yuan, V. Rao, S. Han, and L. Carin, "Hierarchical infinite divisibility for multiscale shrinkage," IEEE Trans. Signal Process. **62**(17), 4363–4374 (2014).

1. Introduction

Range imaging systems, which collect three-dimensional spatial information of the object surface, have a wide range of applications in medical procedures [1], archaeological landscape mapping [2], industrial metrology, 3D printing [3], tracking, and vehicle navigation [4]. Range imaging systems can be roughly categorized into either passive or active sensing methods [5]. Passive sensing methods are cost-effective and simple in implementation, yet are subject to the influences of the environmental lighting condition. Active illumination has more degrees of freedom in controlling the properties of the light source, such as wavelength, polarization, coherence, temporal profile, etc., making this category more versatile in the application space [6,7]. Most active range sensing techniques are based on three major principles: (1) time of flight (TOF), (2) triangulation, and (3) wave diffraction. The central theme of active range imaging is using illumination to encode the depth information into the measurement. TOF imager, capable of long working distance (kilometers) with better resolution than one millimeter, has been widely employed in gaming and automobile industry. Wave diffraction methods, including defocusing point-spread-function engineering [8,9], interferometry, holography [10], etc., have major applications in precision metrology and microscopy, which can achieve sub-wavelength resolution but have a limited working distance. Triangulation methods, calculating the depth information from the illumination angle and the reflection spot location, are suitable for imaging range from 0.1 m to 100 m and can achieve a resolution from 10^{-6} m to 10^{-3} m. Spot/line scanning and Moiré pattern projections are among the most popular triangulation methods.

Recently, inspired by compressive sensing [11], recovery of multiple frames from a single image has been demonstrated [12–15]. In these settings, the measurement is acquired at a low frame rate, with a temporal-spatial coding at the rate faster than the acquisition rate. The high-frame-rate video is computationally recovered subsequently. Instead of using detection-side coding, in [14], a temporally varying structured illumination is introduced, so that the fluorescence signal can be collected by the full aperture of the microscope objective. In this paper, we extend the simple active illumination scheme to implementing both the range encoding and temporal compression simultaneously. We have built an imaging system prototype to validate the imaging principle and have demonstrated a compression ratio of five. The simulated range resolution is in agreement with the resolution calculated from the triangulation methods based on the same system geometry. In this paper, we implement the temporal and range modulation at the same time via structured illumination. High-speed frames and the depth map are recovered via proposed inversion algorithms.

The paper is organized as follows. In Section 2 we develop the mathematical model of the imaging system. Then in Section 3, we propose an iterative reconstruction algorithm that can reconstruct both the reflectance and depth information at a frame rate higher than the acquisition rate. In Section 4, the imaging system prototype is described. Experimental results are presented in Section 5. Section 6 initializes some theoretical discussions on the range resolution and Section 7 concludes the entire paper.

2. Forward model

2.1 Temporal modulation

We consider a high-speed 3D video, modeled as a four-dimensional function $F(x, y, z, t)$. A single snapshot by the camera integrates the high-speed scene $F(x, y, z, t)$ within one frame period T. Figure 1 shows the schematic illustration of the setup. To discern the range of the scene and retrieve the high-speed temporal frames, we use the active illumination $H^*(x, y, z, t)$ to implement encoding in both temporal and range domain. On one hand, the integration time T of the camera limits the bandwidth in the temporal domain. The high-speed

modulation function $H^*(x, y, z, t)$ aliases the high frequency components of scene $F(x, y, z, t)$

into the passband of the measurement, which allows us to reconstruct the high-speed frames. On the other hand, the active illumination varies in the range direction, which provides different structural encoding along z axis.
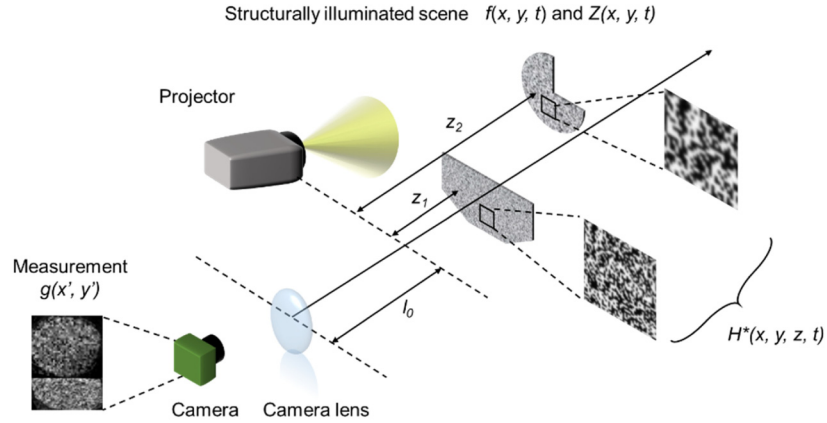

Fig. 1. The forward model of our system. The projector projects high-speed masks onto the scene. The scene is modulated by the projected masks *h*(x, y, t)*, the variants of the original mask at different range locations. The camera integrates the modulated high-speed frames into one measurement.

We use a projector to project high-speed pseudo-random binary masks onto the 3D scene. For a scene composed of opaque objects, *F(x, y, z, t)* can be represented by a reflectance function *f(x, y, t)* and a range function *Z(x, y, t)*. Since we use a single light source, the modulation function *H* (x, y, z, t)* is determined by the original masks *h*(x, y, t),* and the range z. We establish the coordinate system with its origin at the lens of the projector. A snapshot measurement *g(x', y')* is:

$$g(x',y') = \int_{t=0}^{T} f\left(\frac{z+l_0}{f_c}x', \frac{z+l_0}{f_c}y', t\right) \cdot h^*\left(\frac{z+l_0}{f_c}x', \frac{z+l_0}{f_c}y', t\right) dt, \qquad (1)$$

where f_c is the focal length of the camera, l_0 is the distance between projector and camera along z axis.

2.2 Range modulation

One key contribution of our work is to reconstruct the depth information of the scene as well as the temporal super resolution. Different from the previous method, which modulates the range of the scene with varied blur kernels [8], we employ different scales and shift of the masks. Let $h^z(x, y, t)$ denotes the ideal projected images of the original masks $h^0(x, y, t)$ projected to the range z without any objects (e.g., a uniform white background). In other words, h^z is a plane located at range z. Considering the lateral offset d (in x axis) between the camera and the projector as shown in Fig. 2, the ideal projected masks can be expressed as

$$h^z(x,y,t) = h^0\left(\frac{f_p}{z}x + \frac{f_p}{z}d, \frac{f_p}{z}y, t\right), \qquad (2)$$

where f_p is the focal length of the projector. With this structured illumination, the scene at different depths are modulated by masks with various scaling factors f_p/z and the shift $f_p d/z$, which are both the functions of range z. The shift is induced by the lateral offset d. The ideal projected masks h^z can be obtained by calibrations or simulations with the parameters f_p, d, and the origin masks h^0. In other words, the ideal projected masks $h^z(x, y, t)$ are *known as*

priori. This unveils the underlying principle of the range function $Z(x, y, t)$ mentioned in the previous section. Our target is: given $g(x', y')$ with prior of $h^{\bar{z}}$, estimating the reflectance function $f(x, y, t)$ and the range function $Z(x, y, t)$.

3. Reconstruction algorithm

We can discretize high-speed video f, projected masks h^*, and measurements g. Let $\mathbf{F_k} \in R^{N_x \times N_y}$ denote the kth discretized frame of the scene. For each pixel (m, n), $m = 1, \dots, N_x$; $n = 1, \dots, N_y$, we have

$$g_{m,n} = \sum_{k=1}^{N_T} h^*_{m,n,k} \cdot f_{m,n,k}, \tag{3}$$

where we consider N_T discretized high-speed video frames within the integration time T.

Let $\mathbf{f_k}$ be the vectorized form of $\mathbf{F_k}$. The vectorized form of the measurement can be expressed as

$$\mathbf{g} = \mathbf{H}^* \begin{bmatrix} \mathbf{f_1} \\ \vdots \\ \mathbf{f_{N_T}} \end{bmatrix}. \tag{4}$$

The inverse problem can now be formulated as

$$(\hat{\mathbf{f}}, \hat{\mathbf{H}}^*) = \arg_{\mathbf{f}, \mathbf{H}^*} \min \left\| \mathbf{g} - \mathbf{H}^* \mathbf{f} \right\|_2^2 + \tau R(\mathbf{f}). \tag{5}$$

$\mathbf{H}^*$ is the sensing matrix corresponding to the projected masks h^*, and $R(\mathbf{f})$ denotes a regularizer which can be used to impose the sparsity of the signal in the basis such as the wavelet, the discrete cosine transformations (DCT) or the total variation (TV) operator [16]. The regularizer penalizes characteristics of the estimated $\mathbf{f}$ that would result in poor reconstruction. τ is the Lagrange parameter balancing the measurement error and the regularizer. There are two parameters to estimate in Eq. (5), which is non-convex. However, given one, the other one can be solved via existing algorithms [8,14,16]. In the following, we solve the problem in an alternating way by estimating one with the other one fixed.

3.1 Resolving the scene range

Although we cannot directly obtain the projected masks h^*, we know that h^* is a combination of different portions of the ideal masks $h^{\bar{z}}$ at different z. Considering the spatial information as well as the range resolution of the reconstructed results, we divide the measurement into small blocks by imposing local window on the measurement and sliding the window with sub-window step δ in both horizontal and vertical directions to obtain a sequence of blocks [8]. And then we estimate a range $\hat{z}_{block}$ for each block. Let N_z and $\left\{ \mathbf{H^i} \right\}_{i=1}^{N_z}$ denote the number of discretized ranges and the ideal sensing matrix corresponding to the ideal projected masks $h^{\bar{z}}(x, y, t)$ at i^{th} discretized range, respectively. For each block, we can enumerate the ideal sensing matrices $\mathbf{H^i}$ and obtain N_z candidates of the reconstructed fraction of the scene $\mathbf{f_{block}}$ as

$$\hat{\mathbf{f}}^i_{block} = \arg_{\mathbf{f_{block}}} \min \left\| \mathbf{g}_{block} - \mathbf{H}^i \mathbf{f}_{block} \right\|_2^2 + \tau R(\mathbf{f}_{block}), \tag{6}$$

where $i = 1, \dots, N_z$. Equation (6) can be solved by the generally used compressive sensing inversion algorithms, for example, the TV based optimization algorithms [16] or the Bayesian algorithms [17,18]. The second step is to select the one fitting the measurement best,

$$\hat{z}_{block} = z_j, \, j = \arg_i \min \left\| \mathbf{g}_{block} - \mathbf{H}^i \hat{\mathbf{f}}^i_{block} \right\|_2^2, i = 1, \ldots, N_z, \tag{7}$$

where z_j is the jth range. Empirically, we have found adding the regularizer on each block can provide better results

$$j = \arg_i \min \left\| \mathbf{g}_{block} - \mathbf{H}^i \hat{\mathbf{f}}^i_{block} \right\|_2^2 + \tau R(\hat{\mathbf{f}}^i_{block}), i = 1, \ldots, N_z. \tag{8}$$

We assume that the depth map $Z(x, y, t)$ does not change within one measurement, thus we just enumerate N_z possible sensing matrices here. For the scenarios without this assumption, the number of possible sensing matrices is $N_T^{N_z}$ which grows fast with the compression rate N_T. After we select the best range for each block, we can get the corresponding reconstruction for each block. This reconstruction is based on the inferred scale and shift of the original mask h^0 and the measurement g. The final results can be obtained by fusing these reconstruction blocks. However, due to the spatial correlation of different blocks, we refine the results globally via the following step in Section 3.2.

3.2 Refinement of the reconstruction to estimate f

Table 1. Reconstruction algorithm

Require: Measurement **g**, different scales of the mask $\mathbf{H}^i$, i = 1, …, N_z
Divide the measurement into overlapping blocks
For each block, do
Given range z_i, estimate the scene of each block using Eq. (6)
Estimate the range for each block via Eq. (8)
end
Obtain the optimal mask for each pixel
Refine the reconstruction using Eq. (5)

After getting the range for each pixel in Section 3.1, we can obtain the correct mask for each pixel. Equation (4) can now be solved via the video compressive sensing algorithms [12–16], which considers both the global and local information. Table 1 summarizes the proposed inversion algorithm.

4. Experimental implementation

4.1 System setup

The schematic is shown in the Fig. 2. The projector consists of a halogen lamp (Nikon, D-LH), a digital micromirror device (Vialux, DLP4100) and a camera lens (Nikon, 18-55mm). The focal length of the camera lens is 50 mm. A digital micromirror device (DMD) is an array of highly reflective aluminum micromirrors. The maximum frame rate of the DMD is 22.7 kHz which is much faster than the camera. To modulate the high-speed frames of the scene, the projector is used to project pseudo-random binary masks onto the scene at 1000 frames per second (fps) while the camera is acquiring the measurement at 200 fps. Objects at different ranges are modulated by projected masks with various shift and magnifications. The

shift is induced by the separation of the optical axis of the projector and the camera. It is worth noting that the feature size of masks acquired by the camera depends on the distances to the projector L_p and the distance to the camera L_c. Indeed, the pitch is proportional to the ratio of L_p/L_c. If the baseline of the camera and the projector are the same line, the pattern feature on the camera would be the same regardless the range of the objects, in which case the only modulation on the range is the shift. To better discern the pattern at different ranges, we place the camera and the projector at different locations on the z axes. The modulated objects are imaged by a camera lens (Nikon, 18-55mm) onto the camera (JAI, GO5000M). The separation of the axes of the camera and the projector is $d = 135$ mm. The focal length of the camera lens is 50 mm. To ensure the coding process remains time-invariant, we write the pseudo-random patterns into the memory of the DMD prior to the display, and then use an NI board (NI, USB 6353) to synchronize the camera and the DMD in the projector. The NI board generates a pulse to start the DMD display, then generates a 200 Hz square wave to trigger the camera. There is a fixed delay between the DMD and the camera control signals to ensure the synchronization of these two. We use an active area of 296×325 detector pixels to account for the 128×96-pixel pseudo-random patterns with additional zero-padding.
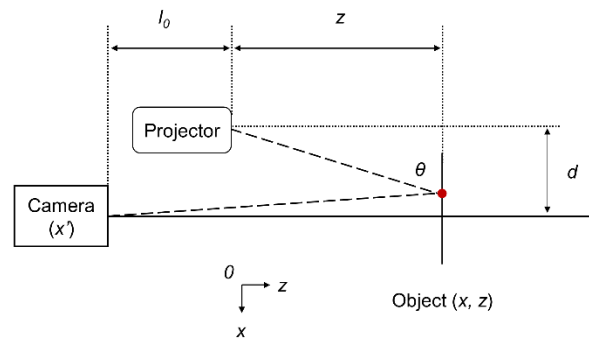

Fig. 2. System geometry. The lateral offset of projector and camera is $d = 135$ mm. The distance between the projector and the camera along the z axis is $l_0 \approx 1.3$ m. θ denotes the projection angle. The focal lengths of the camera and the projector are both 50 mm.

4.2 Calibration

To implement the algorithm, we need to calibrate the sensing matrix $\mathbf{H^i}$ mentioned in Section 3. Firstly, we record pseudo-random mask projected on a white board located at range z_i, $i = 1… N_z$. The increment of the calibrated depths is 10 mm. Secondly, we record the image from an all-on state DMD to correct the nonuniformity of the illumination. Thirdly, we record the image of an all-off state DMD for subtraction of the background. After the corrections to the images of the masks, we can extract $\mathbf{H^i}$. The whiteboard is translated by two motor-driven translation stages (Thorlabs, NRT100 and Newport, LTA-HS) which give us a 150 mm travel range of stable and repeatable translation. Some calibrated masks at different ranges are shown in Fig. 3.
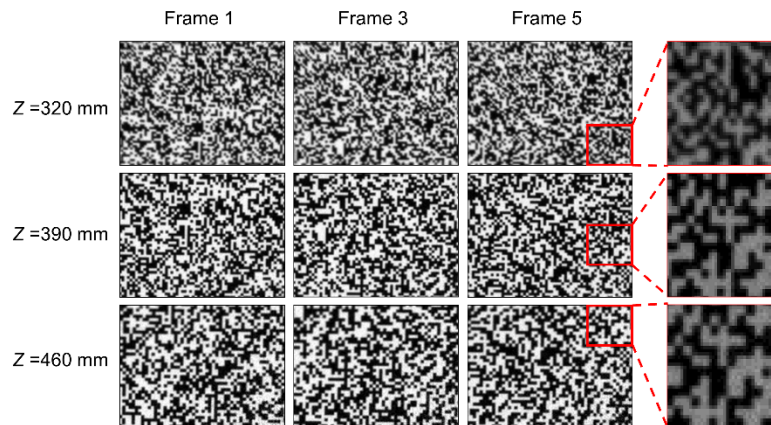
Fig. 3. Calibrated masks at different ranges. Same patterns of the 5th mask projected at different ranges are highlighted with the red boxes. The shift of the red boxes indicates one modulation on the range. The scales of the zoom in patterns is the other modulation on the range.

5. Results

5.1 Simulation results

In this section, we conduct simulations to verify the hardware principle and the proposed inversion algorithms. We use masks with different scales, shown in Fig. 3 as the projected masks. Three ranges are used in our simulation. The resolution of the masks are 160×120, and the scales and shift of the masks are different.

We generate the measurement by the following steps: *i*) generate the video frames, each frame contains objects at all three ranges, *ii*) modulate each frame with corresponding masks with different scales and shift corresponding to the range, *iii*) integrate these modulated frames to one measurement image shown on the top-left of Fig. 4. In the simulation, we set the side length of the blocks to 16 pixels and set the sub-window step δ to 4 pixels. The depths of the letters U, C and F are 320 mm, 390 mm and 460 mm respectively. It can be seen that the scene is composed of three letters at different ranges. The top letter U is the nearest one and moves from left to right. While the bottom letter F is furthest and stationary. The middle letter C moves from right to left. In the simulation, the letters are modulated by three different ranges corresponding to the masks at 320 mm, 390 mm and 460 mm as shown in Fig. 3. After running our algorithm in Section 3, we get the reconstruction results as well as the range for each frame in Fig. 4. It can be observed that the motion of the letters is reconstructed well and the range of the scene is also inferred correctly. Thereby, this simulation verifies the efficacy of the sensing framework and the performance of the proposed reconstruction algorithm.
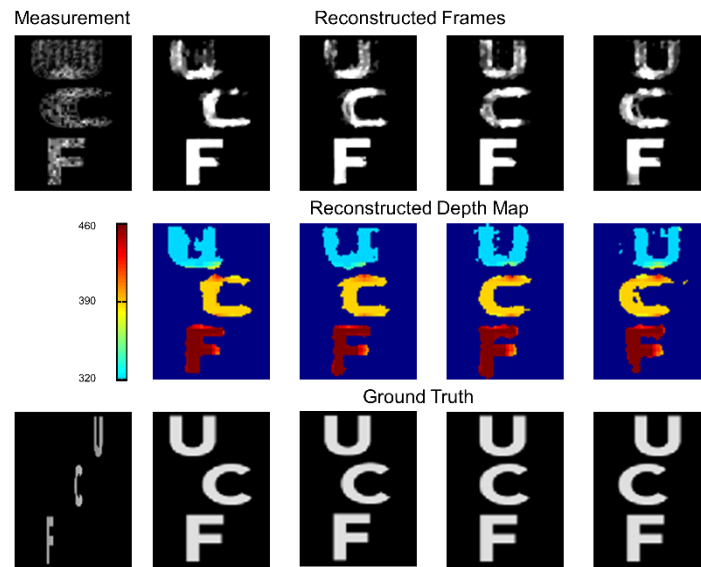
Fig. 4. The measurement and the reconstructed high-speed frames of reflectance are shown in the top row. The corresponding depth maps are shown in the middle row. Bottom row shows the ground truth of the scene.

5.2 Experimental reconstruction

We now use the camera to capture a real high-speed 3D scene and reconstruct it. The setup configuration is shown in Fig. 2. Before acquiring the data, we calibrate the camera with checkerboard to correct the aberrations. We place two objects at different ranges, one is a stationary white board S_1 (68 mm x 28 mm) which is located at $z_1 = 320$ mm away from the projector and $l_0 + z_1 = 1.63$ m away from the camera. The other object is a circular board S_2 (36 mm in diameter) which is fixed on a fan with a notch rotating at 30 rounds per second. A black square tape (18 mm x 18 mm) is added to the surface of S_2, serving as a reference mark. S_2 is placed at 360 mm away from the projector. The camera is operating at 200 fps, and we reconstruct five frames for each measurement. The scene is shown in Fig. 1, and we show one measurement and reconstruction results in Fig. 5. We set the side length of the block to 16 pixels, and set the sub-window step δ to 8 pixels. It can be observed that we can retrieve the scene at 1000 fps and infer the corresponding depth map. A 3D video consisting of 95 reconstructed frames from 19 measurements is included as Visualization 1 in the Supplementary Materials. The computing time for one measurement on our computer with an 8 gigabyte memory and 3.4 GHz processor (Intel, i3-4130) is 48 seconds. As our algorithm is based on blocks and each block is reconstructed independently, parallel computing could be used on a GPU which could significantly accelerate the computing time. There are some artifacts on the boundary of both objects, because the resolution of the depth map depends on the size of the block and the overlapping between the adjacent ones as mentioned in Section 3. Whenever there is a steep slope in the block, the artifacts will appear. These can be improved by using blocks with more overlap with compromised computation time. In addition, the error of the calibration and the noise in the acquisition will deteriorate the reconstruction result, which is inevitable.

Measurement   Reconstructed Frames

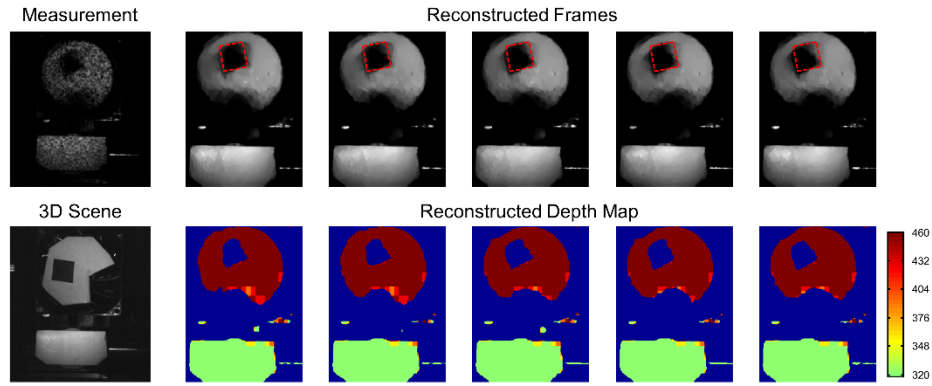3D Scene   Reconstructed Depth Map

Fig. 5. The top row presents a measurement at 200 fps and the reconstructed frames of reflectance at 1000 fps. The object is a disk rotating at 30 rps. The red box indicates the position of the black tape in the first frame. The bottom row shows a photo of the stationary object and the reconstructed depth map. A video of 1000-fps frames has been reconstructed from 200-fps measurement (see Visualization 1).

6. Discussion

We have demonstrated the principle of our prototype, the simulations, and the experimental results. In this section, we discuss the range resolution of our system. For the purpose of simplicity, we reduce our discussion to two-dimensional coordinates (x, z), as depicted in Fig. 2. The focal lengths of the camera and the projector are both 50 mm. The analysis of the range resolution is applicable to 3D situation. The camera coordinates are linked with the object coordinates by similar triangles, $\dfrac{z+l_0}{f_c} = \dfrac{x}{x'}$, where x' is the camera coordinate, f_c is the focal length of the camera lens. Using the system geometry, we can relate the projection angle to the object coordinates, $d - z \cot\theta = x$, where d is the lateral offset of the projector, and θ is the projection angle [7]. Combine the two equations and eliminate the lateral object coordinate, x, we then get the well-known triangulation relation:

$$z = \frac{f_c d - l_0 x'}{x' + f_c \cot\theta}. \tag{9}$$

Equation (9) shows that the range of the reflectance point can be uniquely identified as long as projection angle and the camera coordinates of the reflected point are known. It is worth noting that Eq. (9) differs slightly from the equations in several literature where the camera lens and the projector are placed on the x axis, $(z = 0)$. Let $l_0 = 0$, Eq. (9) will have the same form of the common triangulation equation. The reflectance of the surface is $f(x, z)$, and the measurement is:

$$g(x') = \int t(\theta) f(x, z(x)) \cdot \delta\left(z(x) - \frac{f_c d - l_0 x'}{x' + f_c \cot\theta} \right) d\theta, \tag{10}$$

where $t(\theta)$ is the projection pattern. The Dirac-delta function inside the integral defines the triangulation relation. Despite numerous triangulation methods based on different projection schemes have been proposed, they share the same goal to derive the depth mapping $Z(x)$ and the reflectance of the surface $f(x, z)$ from the measurement $g(x')$. Equation (10) assumes that i) the offset distance d is small relative to the object distance z, ii) the surface range function $Z(x)$ is a convex function so that the projected light reaches all the points on the object surface and the reflected light are not blocked by the surface. Two of the most common choices of the projection patterns are (a) a point or line, which has a projection function $t(\theta) = \delta(\theta - \theta_0)$;

and (b) a periodic pattern (line stripes), whose projection function can be simplified to $t(\theta) = 0.5 + 0.5\cos\left(\dfrac{2\pi}{p}(\theta - \theta_0)\right)$, where p is the period of the stripe pattern. It becomes obvious how we can retrieve the depth information by treating the Eq. (10) as a transform from an illumination pattern $t(\theta)$ to a measurement $g(x')$ from a reflection point at z. For the point projection scenario, the illumination is a delta function. So the measurement $g(x') = g_0\delta(x' - x_0')$ is also a delta function. The detector coordinate is thus determined by the range of the reflector, i.e. $x_0' = \dfrac{(d - z\cot\theta)f_c}{z + l_0} \approx \dfrac{f_c d}{l_0}\left(1 - \left(\dfrac{1}{l_0} + \dfrac{\cot\theta}{d}\right)z\right)$, when $z \ll l_0$.

This equation shows that the shift of the point on the camera depends on the range location of the surface. For single point projection, the range resolution is determined by the localization precision of the reflected point, which is limited by the camera optics and the detector pixel size. For the case where the detection resolution, $\Delta x'$, is limited by the detector pixel size ($\Delta x' = 10$ μm), the range resolution $\Delta z = \dfrac{l_0^2}{f_c d}\Delta x' \approx 3.3mm$.

This formula indicates that the system with a larger offset d could have a higher range resolution. However, due to the consideration on the total field-of-view of the system and the constraints on the system form factor we chose d to be 13.5 cm.

For the scenario where the range profile of the object is piece-wise constant, the phase-shift of a periodic pattern projection can be determined at sub-pixel precision, depending on the number of available fringes. According to the equation of range resolution, the Moiré pattern projection therefore could lead to an improved range resolution, thanks to a finer $\Delta x'$.

In our reconstruction method, the range z is considered as a parameter of the forward model $\mathbf{H^*}$. The reconstruction errors of each block at the different ranges are compared. The range is determined by selecting the one with minimum reconstruction error. Assuming the object is located at range z_1, the resultant measurement is $\mathbf{g} = \mathbf{H^1 f^1}$, and $\hat{\mathbf{f}}^i$ is the optimal reconstruction using the forward matrix $\mathbf{H^i}$ at range z_i, the error function can be expressed as

$$\left\|\mathbf{g} - \mathbf{H}^i\hat{\mathbf{f}}^i\right\|_2^2 = \left\langle \mathbf{g} - \mathbf{H}^i\hat{\mathbf{f}}^i, \mathbf{g} - \mathbf{H}^i\hat{\mathbf{f}}^i \right\rangle = \left\langle \mathbf{g} - \mathbf{H}^i\hat{\mathbf{f}}^i, \mathbf{g} \right\rangle = \left\|\mathbf{g}\right\|_2^2 - \left\langle \mathbf{H}^i\hat{\mathbf{f}}^i, \mathbf{H}^1\hat{\mathbf{f}}^1 \right\rangle, \quad (11)$$

where $<\cdot>$ denotes the inner product of the vectors. Here we impose the fact that the residue of the optimal projection is orthogonal to the projection itself, i.e. $\left\langle \mathbf{g} - \mathbf{H}^i\hat{\mathbf{f}}^i, \mathbf{H}^i\hat{\mathbf{f}}^i \right\rangle = 0$.

Equation (11) shows that if the forward matrix is at the correct range, i.e. $\mathbf{H^1} = \mathbf{H^i}$, the optimal reconstruction error is a minimum. The high range discernibility is equivalent to a large reconstruction error by a small difference in the range, Δz. This speculation is consistent with the analysis from the triangulation equation. For the reflective surface with the same range change, the more significant (shift and magnification) change of the projection pattern, the better the range resolution. To quantify the range resolution, we simulate the projected masks at different ranges and numerically modulate a planar rotational fan with the mask at range 320 mm. Then we reconstruct the high-speed scene using the mask Δz away from 320 mm. The profile of the normalized reconstruction error which is defined in Eq. (8) with the deviation Δz is shown as the blue curve in Fig. 6. In addition, we calculate the correlation of the masks with the deviation Δz. The profile of the correlation subtracted from unity is shown as a red curve in Fig. 6. The simulated correlation curve has an full width half maximum (FWHM) of 3.2 mm which fits the analysis above. The FWHM of the reconstruction error curve is 1.7 mm which shows our system has a potential of finer range resolution. Figure 6 plots the reconstructed 1st frame using the masks with z deviation of 0 mm, 1 mm and 2 mm

from 320 mm, which shows that the range resolution of our system is similar to that of the single point system.
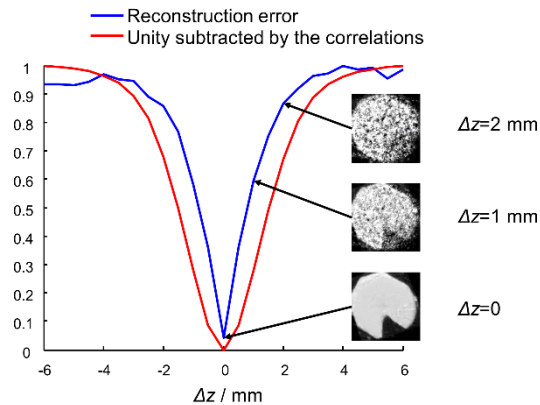

Fig. 6. The simulated range resolution. The red curve is the normalized negative correlation of the masks. The blue curve is the normalized reconstruction error using the masks which are Δz away from the correct one. The inserts indicate the first frame of the reconstructed results using the masks with $\Delta z = 0$, 1 mm, 2 mm.

7. Conclusion

In this work, we describe a novel range imaging system that encodes both range and high-speed temporal information of a scene with a series of binary random patterns projected by a high-speed DMD. We have demonstrated the imaging principle by reconstructing a fast-varying scene that is ~1.7 m away from the camera. To solve the inverse problem, a block-wise alternating algorithm has been developed to reconstruct high-speed temporal and range information. A 1000-fps video of reflectance intensity with depth map is reconstructed from 200-fps measurements. In our reconstruction algorithm, the range resolution is determined by the sensitivity of the reconstruction error, which is inversely related to the correlation of projected patterns at different ranges. The simulation has demonstrated a range resolution better than 3.2 mm, which is in agreement with the range resolution calculated from the triangulation method. To improve the range resolution, a finer projection pattern and further optimization of the system geometry is needed. Our system is simple to implement and has numerous potential applications in high-speed range imaging.