Research Article | Vol. 25, No. 15 | 24 Jul 2017 | OPTICS EXPRESS 18182
Optics EXPRESS

Compressive high-speed stereo imaging

YANGYANG SUN,[1] XIN YUAN,[2,3] AND SHUO PANG[1,4]

[1]College of Optics and Photonics (CREOL), University of Central Florida, 4304 Scorpius Street, Orlando, FL 32816, USA
[2]Nokia Bell Labs, 600 Mountain Avenue, Murray Hill, NJ 07974, USA
[3]xyuan@bell-labs.com
[4]pang@creol.ucf.edu

Abstract: A compressive high-speed stereo imaging system is reported. The system is capable of reconstructing 3D videos at a frame rate 10 times higher than the sampling rate of the imaging sensors. An asymmetric configuration of stereo imaging system has been implemented by including a high-speed spatial modulator in one of the binocular views, and leaving the other view unchanged. We have developed a two-step reconstruction algorithm to recover the irradiance and depth information of the high-speed scene. The experimental results have demonstrated high-speed video reconstruction at 800fps from 80fps measurements. The reported compressive stereo imaging method does not require active illumination, offering a robust yet inexpensive solution to high-speed 3D imaging.

OCIS codes: (330.1400) Vision - binocular and stereopsis; (110.1758) Computational imaging; (150.5670) Range finding.

References and links

1. D. L. Donoho, "Compressed sensing," IEEE Trans. Inf. Theory **52**(4), 1289–1306 (2006).
2. 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).
3. X. Yuan and S. Pang, "Structured illumination temporal compressive microscopy," Biomed. Opt. Express **7**(3), 746–758 (2016).
4. Y. Sun, X. Yuan, and S. Pang, "High-speed compressive range imaging based on active illumination," Opt. Express **24**(20), 22836–22846 (2016).
5. D. Reddy, A. Veeraraghavan, and R. Chellappa, "P2C2: Programmable pixel compressive camera for high speed imaging," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (IEEE, 2011), pp. 329–336.
6. Y. Hitomi, J. Gu, M. Gupta, T. Mitsunaga, and S. K. Nayar, "Video from a single coded exposure photograph using a learned over-complete dictionary," in *Proceedings of the IEEE International Conference on Computer Vision* (IEEE, 2011), pp. 287–294.
7. 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).
8. X. Yuan, "Generalized alternating projection based total variation minimization for compressive sensing," in *International Conference on Image Processing* (IEEE, 2016), pp. 2539 - 2543.
9. J. M. Bioucas-Dias and M. A. Figueiredo, "A New TwIst: Two-Step Iterative Shrinkage/Thresholding Algorithms for Image Restoration," IEEE Trans. Image Process. **16**(12), 2992–3004 (2007).
10. X. Liao, H. Li, and L. Carin, "Generalized alternating projection for weighted-l2,1 minimization with applications to model-based compressive sensing," SIAM J. Imaging Sci. **7**(2), 797–823 (2014).
11. Z. Wang, L. Spinoulas, K. He, L. Tian, O. Cossairt, A. K. Katsaggelos, and H. Chen, "Compressive holographic video," Opt. Express **25**(1), 250–262 (2017).
12. S. Foix, G. Alenyà, and C. Torras, "Lock-in time-of-flight (ToF) cameras: A survey," IEEE Sens. J. **11**(9), 1917–1926 (2011).
13. J. Salvi, S. Fernandez, T. Pribanic, and X. Llado, "A state of the art in structured light patterns for surface profilometry," Pattern Recognit. **43**(8), 2666–2680 (2010).
14. T. E. Bishop and P. Favaro, "The light field camera: Extended depth of field, aliasing, and superresolution," IEEE Trans. Pattern Anal. Mach. Intell. **34**(5), 972–986 (2012).
15. B. Tippetts, D. J. Lee, K. Lillywhite, and J. Archibald, "Review of stereo vision algorithms and their suitability for resource-limited systems," J. Real-Time Image Process. **11**(1), 5–25 (2016).
16. L. I. Rudin and S. Osher, "Total variation based image restoration with free local constraints," in *Proceedings - International Conference on Image Processing* (IEEE, 1994), pp. 31–35.
17. X. Yuan and S. Pang, "Compressive video microscope via structured illumination," in *International Conference on Image Processing* (IEEE, 2016), pp. 1589–1593.

#297014 https://doi.org/10.1364/OE.25.018182
Journal © 2017 Received 31 May 2017; revised 16 Jul 2017; accepted 16 Jul 2017; published 19 Jul 2017

Optics EXPRESS

18. G. Heiko Hirschmüller, "Accurate and Efficient Stereo Processing by Semi-Global Matching and Mutual Information," IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognition (2005), pp. 807–814.
19. V. Kolmogorov and R. Zabih, "Computing Visual Correspondence with Occlusions via Graph Cuts," in *Proc. IEEE International Conference on Computer Vision (ICCV)* (2001), pp. 508–515.
20. Z. Zhang, "A flexible new technique for camera calibration," IEEE Trans. Pattern Anal. Mach. Intell. **22**(11), 1330–1334 (2000).
21. Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," IEEE Trans. Pattern Anal. Mach. Intell. **26**(9), 1124–1137 (2004).
22. V. Kolmogorov and R. Zabih, "What energy functions can be minimized via graph cuts?" IEEE Trans. Pattern Anal. Mach. Intell. **26**(2), 147–159 (2004).
23. F. Gamboa and E. Gassiat, "Bayesian methods and maximum entropy for ill-posed inverse problems," Ann. Stat. **25**(1), 328–350 (1997).
24. X. Yuan, Y. Sun, and S. Pang, "Compressive video sensing with side information," Appl. Opt. **56**(10), 2697–2704 (2017).
25. X. Yuan, Y. Sun, and S. Pang, "Video compressed imaging using side information," Proc. SPIE **10222**, 102220I (2017).
26. L. Zhu, Y. Chen, J. Liang, Q. Xu, L. Gao, C. Ma, and L. V. Wang, "Space- and intensity-constrained reconstruction for compressed ultrafast photography," Optica **3**(7), 694–697 (2016).
27. J. Liang, C. Ma, L. Zhu, Y. Chen, L. Gao, and L. V. Wang, "Single-shot real-time video recording of a photonic Mach cone induced by a scattered light pulse," Sci. Adv. **3**(1), e1601814 (2017).
28. X. Yuan, X. Liao, P. Llull, D. Brady, and L. Carin, "Efficient patch-based approach for compressive depth imaging," Appl. Opt. **55**(27), 7556–7564 (2016).

1. Introduction

High-speed depth sensing plays an important role in interactive gaming systems, autonomous driving vehicles and augmented/virtual reality devices. Similar to common imaging systems, the frame-rate of such systems is limited by the read-out speed of the sensor, which is on the order of tens frames per second (fps). Inspired by compressive sensing [1], several computational imaging systems have been developed to go beyond the frame-rate limit of the sensor [2–6]. In these systems, the high-speed scenes are spatially modulated at a temporal frequency higher than the band limit of the sensor. The high-frame-rate scene can be reconstructed from the encoded measurement [7–10]. More recently, this effort has been extended to compressive 3D video sensing. Two noticeable examples are demonstrated based on structured illumination [4] and hologram [11]. Both approaches rely on active illumination, which not only increases the cost of the system but also makes it vulnerable to the interference from the environment [12–15]. Passive imaging system would provide a robust and inexpensive solution, making high-speed 3D imaging more accessible. In this work, we demonstrate a compressive stereo imaging setup as an attempt to implement a passive high-speed depth sensing system.

Bearing the system cost in mind, we engineered an *asymmetric* stereo imaging system that includes the high-speed modulator in only one of the optical paths, while keeping the other optical path unmodified, simply a low-frame-rate camera to capture a low-frame-rate blurry scene. To reconstruct the high-speed 3D scene, a general framework is proposed to estimate the depth and intensity information from the two measurements. The major challenge is to estimate the depth from the two asymmetric optical paths and in order the address this, we develop a two-step algorithm, in which the first step recovers the high-speed scene from the modulated optical path and the second step extracts the depth of the scene by employing the information from both measurements.

In the following, we first describe the setup and the mathematic model of our system, and then explain the reconstruction algorithm. A prototype based on digital micromirror device (DMD) is presented. We further demonstrate the imaging performance of the prototype. Finally, we will discuss the limits and the outlook of the compressive stereo imaging system.

2. Theory

Stereo imaging system estimates 3D scene from measurements taken from left and right views. Two pixels in these measurements are said to correspond if they refer to the same

Optics EXPRESS

element in the scene. In rectified epipolar geometry, corresponding pixels are on the same row, and the location difference of these pixels is called disparity. The depth of an object in the scene can be inferred from the disparity between these two measurements.

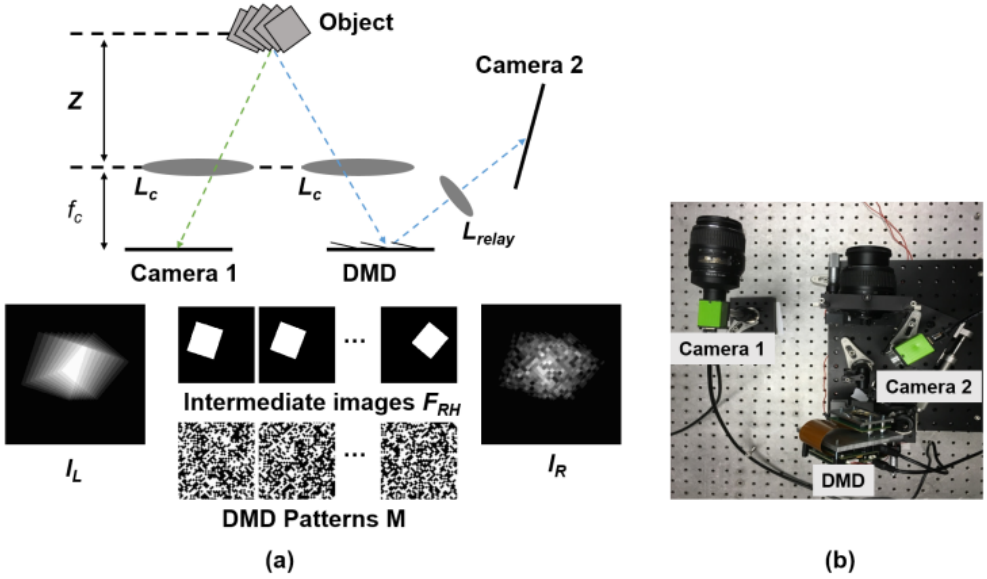

Fig. 1. System schematic (a) On the left-view optical path, Camera 1 records low-speed measurement I_L. On the right-view optical path, high-speed right-view scene F_{RH} are encoded by the DMD with N distinct patterns M. The coded right-view scene is then relayed by lens L_{relay} and recorded by Camera 2 within the exposure time to form the right-view measurement I_R. (b) A photo of the setup.

2.1 Imaging forward model

Figure 1 depicts our compressive stereo imaging system. Both left-view and right-view measurements are synchronized and sampled at a low frame-rate. The left-view measurement I_L captures the summation of the high-speed scene within exposure time. On the right-view optical path, the high-speed scene F_{RH} is modulated by N high-speed pseudo-random patterns M during the exposure time. The modulated scene is then relayed to Camera 2 forming the right-view measurement I_R. Considering the stereo measurements of each view has $N_x \times N_y$ pixels, the $(i,j)^{th}$ pixel can be expressed as:

$$I_R(i,j) = \sum_{n=1}^{N} F_{RH}(i,j,n) M(i,j,n), \tag{1}$$

$$I_L(i,j) = \sum_{n=1}^{N} F_{LH}(i,j,n), \tag{2}$$

where $i = 1,\ldots,N_x, j = 1,\ldots,N_y$, F_{LH}, F_{RH} are the left- and right- view high-speed scene, respectively. As mentioned above, in the stereo imaging system, the high-speed depth information lies in the correspondence between F_{RH} and F_{LH}. However, neither F_{RH} nor F_{LH} can we measure directly since the frame-rate of F_{RH} and F_{LH} exceeds that of the cameras, let alone the correspondence. Our contribution is to estimate the high-speed 3D scene from low frame-rate measurements I_L and I_R.

Optics EXPRESS

2.2 Reconstruction

After capturing the measurements shown in Eqs. (1) and (2), we aim to estimate the high-speed scene as well as the depth. Let $F(i,j,k,n)$ denote the high-speed 3D scene that we are interested, where *i, j* symbolize the spatial indices, k signifies the depth information and n is the high-speed frame index. The relation between the left- and right-view high-speed scenes can be expressed as $F_{LH} = HF_{RH}$ where H is a transformation matrix depending on the depth. Since we only need to estimate F_{RH} and H to obtain the high-speed 3D scene, the reconstruction problem can be formulated as

$$(\hat{F}_{RH}, \hat{H}) = \arg\min_{F_{RH},H} \left\| I_R - \sum_{n=1}^{N} F_{RH} \cdot M \right\| + \left\| I_L - \sum_{n=1}^{N} HF_{RH} \right\|, \quad (3)$$

Unfortunately, Eq. (3) is ill-posed and cannot be solved directly. Thanks to the recent advances in compressive sensing, we can add prior knowledge on F_{RH} and H to make Eq. (3) feasible to solve. This leads to

$$(\hat{F}_{RH}, \hat{H}) = \arg\min_{F_{RH},H} \left\| I_R - \sum_{n=1}^{N} F_{RH} \cdot M \right\| + \lambda\Phi(F_{RH}) + \left\| I_L - \sum_{n=1}^{N} HF_{RH} \right\| + \kappa\Omega(H), \quad (4)$$

where Φ and λ are the regularizer and weight for F_{RH}, Ω and κ are the regularizer and weight for H respectively. The two arguments are coupled through the third term in Eq. (4). In this paper, we ignore the impact of the coupled term in estimating F_{RH} and propose a two-step algorithm to solve the high-speed video (F_{RH}) and estimate the high-speed depth maps (H). We estimate F_{RH} with first two terms in Eq. (4), and then estimate H using the last two terms. This approximation works well in practice as our results shown. In the following, we consider F_{LH}, F_{RH} in the rectified epipolar geometry, and therefore H can be explicitly represented by the disparity shown in Eq. (6).

3. Algorithm

We propose a two-step algorithm demonstrated in Fig. 2 to address the challenging problem in Eq. (4). The right-view high-speed scene F_{RH} is first reconstructed from the snapshot I_R. Secondly, we solve the correspondence problem, i.e, H in Eq. (4), between single left-view measurement I_L and N-frame high-speed right-view scene F_{RH}.

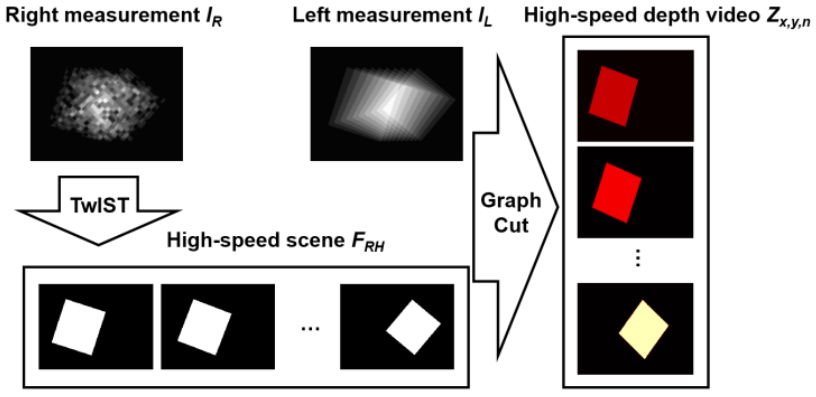

Fig. 2. The flow chart of the reconstruction algorithm. The high-speed scene F_{RH} is reconstructed from the modulated measurement I_R using video compressive sensing inversion algorithm, TwIST. Then, the high-speed depth maps are estimated from I_L and F_{RH} by our one-to-N correspondence algorithm based on Graph Cut.

Optics EXPRESS

3.1 Estimate the high-speed scene

In the first step, we reconstruct high-speed scene F_{RH} from the snapshot I_R in Eq. (1). This is a video compressive sensing inversion problem. Since the DMD multiplexes N frames and collapses into one measurement, this inversion problem is ill-posed. By exploiting the underlying structure within the data, modern reconstruction algorithms [7–10] can reconstruct the data with different degrees of success. Here, we use the iterative reconstruction algorithm TwIST to solve the optimization problem [9]

$$\hat{F}_{RH} = \arg\min_{F_{RH}} \left\| I_R - \sum_{n=1}^{N} F_{RH} \cdot M \right\| + \lambda \Phi(F_{RH}), \tag{5}$$

where $\Phi()$ is the regularizer. The total variation regularizer is employed to promote piece-wise smoothness, since natural scenes are usually sparse in spatial gradients [16,17]. After this step, we obtain N high-speed frames F_{RH} from right-view optical path, while only a single blurry measurement I_L is available from the left-view optical path.

3.2 Infer the depth

The second step is to estimate correspondence between I_L and F_{RH}. Although various correspondence algorithms [18,19] exist to estimate the disparity map, they aim to find the one-to-one correspondence between two measurements. By contrast, in our system, I_L does not correspond to any single frame in F_{RH} but to all N high-speed frames. To explicitly represent H in Eq. (4), let $D(i, j, n)$ denote the disparity of $(i, j)^{th}$ pixel in left-view *low-speed measurement* I_L and n^{th} frame in the right-view *high-speed scene* F_{RH}. The un-occluded pixels in I_L can be represented as

$$I_L(i,j) = \sum_{n=1}^{N} F_{RH}\left(i - D(i,j,n), j, n\right), \tag{6}$$

The depth $Z(i, j, n)$ can be calculated by

$$Z(i,j,n) = \frac{f_c b}{D(i,j,n)}, \tag{7}$$

where f_c is the focal length of the camera lens, b is the baseline length; f_c and b can be obtained by calibration [20]. Our challenge is to compute the one-to-N correspondence between I_L and N-frame video F_{RH} thus the vital ingredient of our algorithm. We formulate this as an energy minimization problem, and propose a correspondence algorithm based on Graph Cut [19,21]. More specifically, we estimate the disparity by

$$\hat{D} = \arg\min_{D} E_{data}(D) + E_{regularizer}(D), \tag{8}$$

where E_{data} and $E_{regularizer}$ denote the data term and the regularization term of the energy function, respectively. In our system, $E_{data}(D)$ is used to measure the similarity of the corresponding pixels according to Eq. (6). In the rectified measurements, corresponding pixels of different perspectives are in the same row. Employing the absolute difference as metric, $E_{data}(D)$ is defined by a one-to-N assignment

$$E_{data}(D) = \sum_{i,j} \left| I_L(i,j) - \sum_{n} F_{RH}\left(i - D(i,j,n), j, n\right) \right|. \tag{9}$$

To engineer diverse problems during the matching process in stereo imaging systems, the regularization term $E_{regularizer}$ is composed of three terms:

Optics EXPRESS

$$E_{regularizer} = E_{occlusion} + E_{uniqueness} + E_{smoothness} \tag{10}$$

$E_{occlusion}$ is a penalty to the occluded pixels,

$$E_{occlusion} = K_{occ} N_{occ}, \tag{11}$$

where K_{occ} is a tuning parameter and N_{occ} is the number of matching pairs labeled as occluded. Intuitively, if the data term (absolute difference) of a matching is larger than K_{occ}, we tend to set this matching as occluded. Smaller value of K_{occ} leads to more occlusions in D. The approach of selecting K_{occ} is stated in [19].

$E_{uniqueness}$ enforces the uniqueness of the matching D,

$$E_{uniqueness} = K_{uniqueness} T(D), \tag{12}$$

where $K_{uniqueness}$ is a corresponding constant far larger than other terms in Eq. (10), $T(D)$ quantifies the uniqueness of D. $T(D) = 1$ if any pixel in I_L or F_{RH} is involved in more than one assignments, otherwise $T(D) = 0$. The uniqueness term discards the non-unique estimations of D. $E_{smoothness}$ promotes the piece-wise smoothness in D. For two adjacent pixels $\{p, s\}$ in I_L, if p and s have different disparities, we give a penalty V. Let D_p and D_s denote the disparity at pixel p and s in I_L, the smoothness term is

$$E_{smoothness} = \sum_{s,p \in \Pi, D_s \neq D_p} V, \tag{13}$$

where $\prod$ is a neighborhood system and V takes binary value. Specifically, if $|I_L(p) - I_L(s)|$ is above a pre-defined threshold η, $V = V_1$, otherwise $V = V_2$, where $V_1 < V_2$. In our experiment, we use $V_1 = 0.5V_2$, η is 0.6 of the maximum intensity in I_L. The threshold η is a level of intensity jump in I_L. The underlying rationale is to match the depth jump with the intensity jump. By using these graph representable energy terms and an appropriate definition of the smoothness term, we can find a strong local minimum of the problem in Eq. (8) via Graph Cut [21–23]. Empirically, we have found that this definition has led to a strong local minimum of the problem in Eq. (10), which is sufficient for our applications.

3.3 Derive the depth resolution

The depth resolution can be derived from Eq. (7). By taking derivative of the depth, Z, regarding to the disparity, D, the depth resolution is $dZ = \frac{f_c b}{D^2} dD$. Substitute D with $\frac{f_c b}{Z}$ in the denominator, the depth resolution can be expressed as

$$dZ = \frac{Z^2}{f_c b} dD, \tag{14}$$

where dD is the spatial resolution of the images. In our setup, f_c = 26 mm, b = 174 mm, the scene is around 1.6 m away from cameras. The pixel size of the camera dD = 5.5 μm. However, the spatial resolution of reconstructed high-speed video also depends on the feature size of modulation patterns which is 41.1 μm. In our experiments, for computational efficiency, we down-sampled reconstructed F_{RH} and left-view measurement I_L by 8, and estimate the high-speed depth map based on the down-sampled images. Therefore, the theoretical depth resolution in our experimental results is around 2.4 cm.

Optics EXPRESS

4. Experimental setup

We built our prototype demonstrated in Fig. 1(b). The same camera lens (Nikon, 18-55mm) and camera (JAI, GO 5000M) are used on both optical paths. The cameras are triggered and synchronized by the data acquisition board (NI, USB6353). On the left-view optical path, the camera is placed on the back focal plane of the camera lens. On the right-view optical path, the high-speed scene is modulated by a DMD (Vialux, DLP7000), and then relayed to the camera by the relay lens (Edmund Optics, 30mm, f/8). The pitch of the DMD is 13.7 μm with fill factor of 0.92. The DMD is working in 3 × 3 binning mode. To extend the working distance, we utilize another relay lens to relay the intermediate images (on the back focal plane of the camera lens) to the DMD.

5. Results and discussion

We captured high-speed scenes using our setup to demonstrate the capability of our proposed system as well as the algorithm. We first tried a challenging object which was moving in *z*-direction. The scene consists of a stationary book and a ball moving away from the camera. The second example has complicated motion and ultrafast moving objects.

5.1 High-speed depth variant scene

In the first example, the cameras were operating on 30fps. The compression ratio N equals to 10. The left and right measurements are shown in Fig. 3(a). The different directions of motion blurs on I_L and I_R indicate the varying depth of the ball within the exposure time. This is a challenging problem for any existing stereo imaging systems and correspondence algorithms: estimating the varying depth from motion blur without the knowledge of the shape of object. By contrast, we address this using our proposed reconstruction framework.

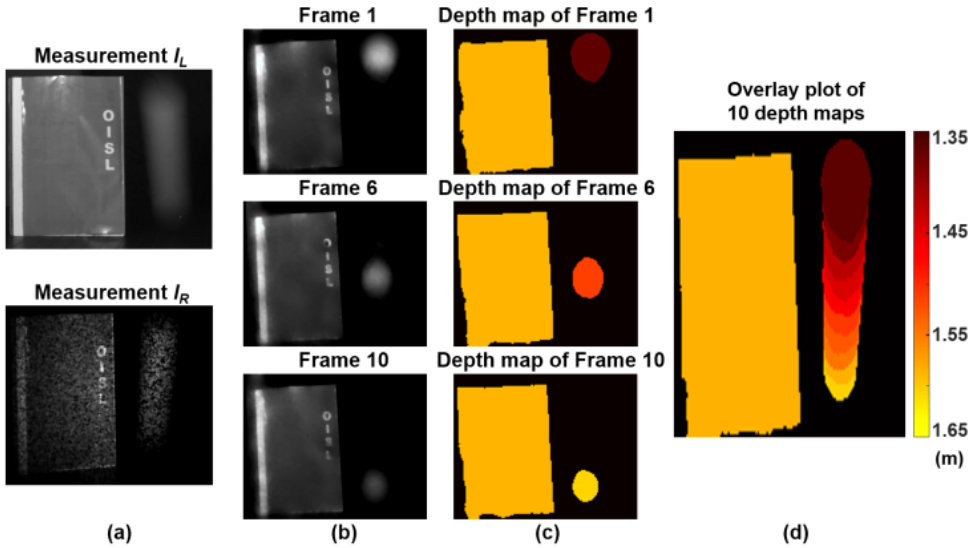

Fig. 3. Reconstruction of a backward-moving ball. (a) Measurements from two optical paths in our system. The different traces of the motion blurs indicate the varying depth of the moving ball. (b) 1^{th}, 6^{th} and 10^{th} fames of reconstructed high-speed video from single measurement I_R. (c) 1^{th}, 6^{th} and 10^{th} fames of reconstructed high-speed depth map. (d) An overlay plot of 10 depth maps within a single exposure. The gradient of the color implies that the ball is moving away from our imaging system. (See Visualization 1).

Following the flow-chart of our algorithm, we first reconstruct 10 high-speed frames from the right measurement I_R and the calibrated DMD patterns M, with results shown in Fig. 3(b). After this, we send these 10 frames along with the left measurement to the correspondence

Optics EXPRESS

algorithm we have built in Eq. (10). The outputs of the algorithm are 10-frame depths as shown in Fig. 3(c). Considering N_x columns in the measurement, $(N_x)^N$ different disparities are possible in one-to-N matching while only N_x possible disparities in one-to-one matching. The size of the searching space will be a challenge when N and N_x become large. In this example, N_x = 125 after down-sampling. Under the linear motion assumption along z axis in a short duration, we can decrease the searching space size from $(N_x)^N$ to N_xN_s, where N_s is the number of possible velocities along z axis that can be detected. The frame-rate of the reconstructed video is 300fps which is 10 times as that of the camera. The intensity difference of the reconstructed ball in different frames indicates that the illumination power of ball is getting weaker when it is getting away from the camera. The overlay plot of the depth map in Fig. 3(d) demonstrates the estimated motion of the ball. The depth increment of adjacent frames is around 23 mm which corresponds to the theoretical disparity of one pixel, i.e., around 2.4 cm. The estimated average velocity of the ball along z axis is 6.7 m/s. The book is located 1556 mm and the estimated depth for the book is 156 cm. The error is within the depth resolution which matches the theoretical analysis in Sec 3.3. This example clearly demonstrates that the shape variation problem is mitigated by our algorithm, and the z-direction resolution of our imaging system is on the order of centimeter. The measurements are cropped into 1000 × 1200 in pixels and the computation time is 23 minutes with a 4.2G Hz CPU (Intel i7 – 7700K). It could be accelerated with alternative parallel algorithm [7] for TwIST and parallel strategy for Graph Cut.

5.2 High-speed complicated motion scene

In the second example, we test our system with scenes containing more complicated motion by the cameras operated at a higher frame rate. The scene consists of a stationary box with letters "UCF", a fast-moving triangular shuriken and a moving rectangular shuriken. The camera is operating at 80fps, while the compression ratio is still 10. Thus, the expected frame-rate of the reconstructed video is 800fps.

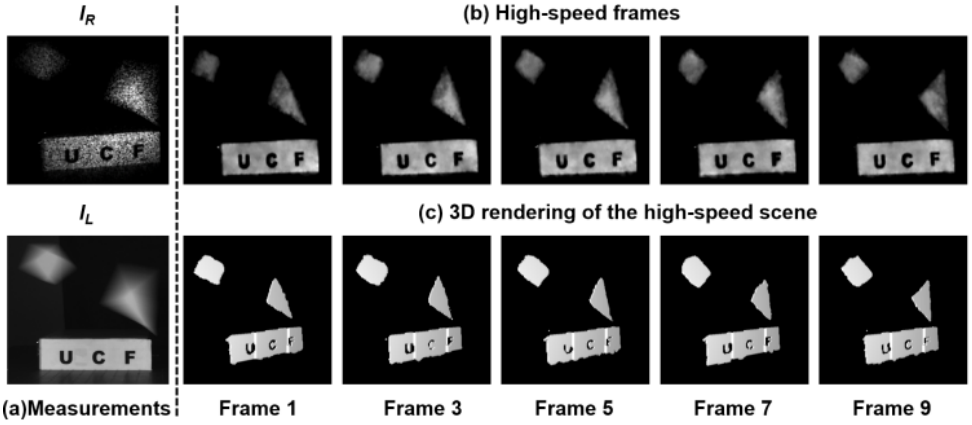

Fig. 4. Reconstruction of an 800fps high-speed scene with two flying "shurikens". (a) Stereo imaging measurements. (b) Selected frames of reconstructed 800fps video. (c) High-speed 3D scene. The rectangular shuriken rotated about 30 degrees within the exposure time, and the triangular shuriken rotated about 20 degrees.

As shown in Fig. 4, the motion blur in the measurement indicates that the motions of the shurikens are mixture of transformation and rotation. Similar to the first example, we first reconstruct the high-speed video frames, now at 800fps, shown in the top-right of Fig. 4. Then our correspondence algorithm provides the depth maps for these 10 frames. An 800fps 3D video is reconstructed from 80fps measurements (See Visualization 2). The rectangular shuriken rotated 30 degrees while the triangular shuriken rotates 20 degrees within the

Optics EXPRESS

exposure time. The 3D rendering plots of the scene are shown in Fig. 4(c). The "UCF" stationary box is not flat relative to the right camera, but has a slope in depth, from 149cm to 153cm. It is encouraging that our algorithm can resolve this depth slope as it gives three steps for this box (Fig. 4(c)), showing that our system is capable of resolving the depth at centimeter level. This is consistent with the observation in the first example and the theoretical analysis in Sec. 3.3. The depths of the triangular and rectangular shurikens are 172cm and 158cm, respectively. The depth estimation matched our setup.

6. Summary

In summary, we have reported a high-speed compressive stereo imaging system, and a two-step inverse algorithm. We have reconstructed a 3D video at 800 fps from coded stereo measurements at 80 fps. Our system exploits the correlations of temporal, spatial and depth channels of the information in passive depth sensing.

From the hardware perspective, the temporal limit of our system is the DMD refreshing rate. A faster modulation can lead to an even faster frame rate. However, this does not mean the frame rate can be increased in this fashion. The faster frame rate will also reduce the signal to noise ratio (SNR) of the measurement, which is detrimental to the reconstructed images. The depth resolution of system is limited by the triangulation geometry of the stereo imaging system, which is on the order of centimeter. Specifically, the depth resolution depends on the spatial resolution of the images, the distance between the scene and the camera, the focal length, and the baseline distance. In the compressive sensing system, the spatial resolution is also affected by the matching between the pixel size of the camera and the feature size of modulation pattern on DMD. On one hand, large feature size would result in low spatial resolution and larger errors in depth triangulation; on the other hand, smaller feature size could result in a poor calibration, which would inversely affect the reconstruction.

Here we would like to mention the option of using two spatial modulators, e.g. two DMDs, in both optical paths. The reconstruction can be simply divided to 1) recovering the high-speed videos from both paths and 2) calculating the corresponding depth maps. In addition to the obvious advantages of lower power consumption and lower system cost of our imaging setup, leaving one optical path unmodified maximizes the light collection efficiency of the stereo imaging system. Our recent results show that this light-collection improvement could lead to a superior reconstruction in the temporal compressive system with two identical channels [24–27].

We envision an integrated reconstruction frame work that merges the current two reconstruction steps. The depth estimation could be used to transform the left-view measurement as side information to improve the high-speed reconstruction of the right-view [24]. An iterative process of updating the right-view reconstruction and depth map could thus be implemented [28]. Different from active illumination system which is sensitive to the ambient light, the reported method is suitable for passive depth sensing system and can be directly implemented using a color camera, making the RGBD sensing system more accessible.