Research Article | Vol. 25, No. 1 | 9 Jan 2017 | OPTICS EXPRESS 250

Optics EXPRESS

Compressive holographic video

ZIHAO WANG,[1,*] LEONIDAS SPINOULAS,[1] KUAN HE,[1] LEI TIAN,[2] OLIVER COSSAIRT,[1] AGGELOS K. KATSAGGELOS,[1] AND HUAIJIN CHEN[3]

[1]*Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL 60208, USA*
[2]*Department of Electrical and Computer Engineering, Boston University, Boston, MA 02215, USA*
[3]*Department of Electrical and Computer Engineering, Rice University, Houston, TX 77005,USA*
[*]*zwinswang@gmail.com*

Abstract: Compressed sensing has been discussed separately in spatial and temporal domains. Compressive holography has been introduced as a method that allows 3D tomographic reconstruction at different depths from a single 2D image. Coded exposure is a temporal compressed sensing method for high speed video acquisition. In this work, we combine compressive holography and coded exposure techniques and extend the discussion to 4D reconstruction in space and time from one coded captured image. In our prototype, digital in-line holography was used for imaging macroscopic, fast moving objects. The pixel-wise temporal modulation was implemented by a digital micromirror device. In this paper we demonstrate 10× temporal super resolution with multiple depths recovery from a single image. Two examples are presented for the purpose of recording subtle vibrations and tracking small particles within 5 ms.

© 2017 Optical Society of America

OCIS codes: (090.1995) Digital holography; (110.1758) Computational imaging.

References and links

1. E. Candès and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," IEEE Trans. Inform. Theory **52**(12) 489-509 (2006).
2. E. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," IEEE Trans. Inform. Theory **52**(12) 5406-5425 (2006).
3. D. L. Donoho, "Compressed sensing," IEEE Trans. Inform. Theory **52**(4) 1289–1306 (2006).
4. M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, "Compressed sensing MRI" IEEE Signal Processing Magazine **25**(2), 72–82 (2008).
5. L. Gan, "Block compressed sensing of natural images," in *International Conference on Digital Signal Processing* (2007), pp. 403–406.
6. D. Brady, K. Choi, D. Marks, R. Horisaki, and S. Lim, "Compressive holography" Opt. Express **17**(15), 13040–13049 (2009)
7. D. Gabor, "A new microscopic principle," Nature **161**(4098), 777–778 (1948).
8. P. Memmolo, L. Miccio, M. Paturzo, G. Di Caprio, G. Coppola, P. A. Netti, and P. Ferraro, "Recent advances in holographic 3D particle tracking," Adv. Opt. Photon. **7**(4), 713–755 (2015).
9. L. Xu, X. Peng, J. Miao, and A. K. Asundi, "Studies of digital microscopic holography with applications to microstructure testing," Appl. Opt. **40**(28), 5046–5051 (2001).
10. T. Su, L. Xue, and A. Ozcan, "High-throughput lensfree 3D tracking of human sperms reveals rare statistics of helical trajectories," Proc. Natl. Acad. Sci. U. S. A. **109**(40), 16018–16022 (2012).
11. Q. Lü, Y. Chen, R. Yuan, B. Ge, Y. Gao, and Y. Zhang, "Trajectory and velocity measurement of a particle in spray by digital holography," Appl. Opt. **48**, 7000–7007 (2009).
12. L. Dixon, F. C. Cheong, and D. G. Grier, "Holographic deconvolution microscopy for high-resolution particle tracking," Opt. Express **19**, 16410–16417 (2011).
13. M. J. Saxton and K. Jacobson, "Single-particle tracking: applications to membrane dynamics," Ann. Rev. Biophys. Biomolecular Structure **26**(1), 373–399 (1997).
14. J. Katz and J. Sheng, "Applications of holography in fluid mechanics and particle dynamics," Ann. Rev. Fluid Mech. **42**, 531–555 (2010).
15. L. Tian, N. Loomis, J. A. Domínguez-Caballero, and G. Barbastathis, "Quantitative measurement of size and three-dimensional position of fast-moving bubbles in air-water mixture flows using digital holography," Appl. Opt. **49**(9), 1549–1554 (2010).
16. W. Xu, M. H. Jericho, H. J. Kreuzer, and I. A. Meinertzhagen, "Tracking particles in four dimensions with in-line holographic microscopy," Opt. Lett. **28**(3), 164–166 (2003).

#279623 http://dx.doi.org/10.1364/OE.25.000250
Journal © 2017 Received 28 Oct 2016; revised 15 Dec 2016; accepted 15 Dec 2016; published 4 Jan 2017

Optics EXPRESS

17. B. J. Nilsson and T. E. Carlsson, "Simultaneous measurement of shape and deformation using digital light-in-flight recording by holography," Opt. Eng. **39**, 244–253 (2000).
18. M. K. Kim, *Digital Holographic Microscopy* (Springer, 2011).
19. J. Garcia-Sucerquia, W. Xu, S. K. Jericho, P. Klages, M. H. Jericho, and H. J. Kreuzer, "Digital in-line holographic microscopy" Appl. Opt. **45**(5), 836–850 (2006).
20. W. Xu, M. H. Jericho, I. A. Meinertzhagen, and H. J. Kreuzer, "Digital in-line holography for biological applications," Proc. Natl. Acad. Sci. **98**(20), 11301–11305 (2001).
21. W. Chen, L. Tian, S. Rehman, Z. Zhang, H. P. Lee, and G. Barbastathis, "Empirical concentration bounds for compressive holographic bubble imaging based on a Mie scattering model" Opt. Express 23(4), 4715-4725 (2015).
22. X. Yu, J. Hong, C. Liu, and M. K. Kim, "Review of digital holographic microscopy for three-dimensional profiling and tracking," Opt. Eng. **53**(11), 112306–112306 (2014).
23. N. Salah, G. Godard, D. Lebrun, P. Paranthoën, D. Allano and S. Coëtmellec, "Application of multiple exposure digital in-line holography to particle tracking in a Bénard–von Kármán vortex flow," Meas. Sci. Technol. **19**(7), 074001 (2008).
24. X. Yu, J. Hong, C. Liu, and M. K. Kim, "Review of digital holographic microscopy for three dimensional profiling and tracking," Opt. Eng. **53**, 112306 (2014).
25. S. Lim, D. L. Marks, and D. J. Brady, "Sampling and processing for compressive holography," Appl. Opt. **50**, H75–H86 (2011).
26. Y. Rivenson, A. Stern, and B. Javidi, "Compressive Fresnel holography," J. Disp. Technol. **6**(10), 506–509 (2010).
27. Y. Liu, L. Tian, J. W. Lee, H. Y. Huang, M. S. Triantafyllou, and G. Barbastathis, "Scanning-free compressive holography for object localization with subpixel accuracy," Opt. Lett. **37**(16), 3357–3359 (2012).
28. Y. Liu, L. Tian, C. Hsieh, and G. Barbastathis, "Compressive holographic two-dimensional localization with $1/30^2$ subpixel accuracy," Opt. Express 22, 9774-9782 (2014).
29. J. Song, C. L. Swisher, H. Im, S. Jeong, D. Pathania, Y. Iwamoto, M. Pivovarov, R. Weissleder, and H. Lee, "Sparsity-based pixel super resolution for lens-free digital in-line holography," Sci. Rep. **6**, (2016).
30. J. Hahn, S. Lim, K. Choi, R. Horisaki, and D. J. Brady, "Video-rate compressive holographic microscopic tomography," Opt. Express **19**, 7289–7298 (2011).
31. M. M. Marim, M. Atlan, E. Angelini, and J.-C. Olivo-Martin, "Compressed sensing with off-axis frequency-shifting holography," Opt. Lett. **35**, 871–873 (2010).
32. C. F. Cull, D. A. Wikner, J. N. Mait, M. Mattheiss, and D. J. Brady, "Millimeter-wave compressive holography," Appl. Opt. **49**, E67–E82 (2010).
33. R. Horisaki, Y. Ogura, M. Aino, and J. Tanida, "Single-shot phase imaging with a coded aperture," Opt. Lett. **39**(22), 6466–6469 (2014).
34. R. Egami, R. Horisaki, L. Tian, and J. Tanida, "Relaxation of mask design for single-shot phase imaging with a coded aperture," Appl. Opt. **55**(8), 1830–1837 (2016).
35. R. Raskar, A. Agrawal, and J. Tumblin, "Coded exposure photography: motion deblurring using fluttered shutter," ACM Trans. Graphics (TOG) **25**(3), 795–804 (2006).
36. G. Bub, M. Tecza, M. Helmes, P. Lee, and P. Kohl, "Temporal pixel multiplexing for simultaneous high-speed, high-resolution imaging," Nat. Methods **7**, 209–211 (2010).
37. M. Gupta, A. Agrawal, A. Veeraraghavan, and S. G. Narasimhan, "Flexible voxels for motion-aware videography," in *European Conference on Computer Vision* (2010), pp. 100–114.
38. D. Reddy, A. Veeraraghavan, and R. Chellappa, "P2C2: Programmable pixel compressive camera for high speed imaging," in *IEEE Conference on Computer Vision and Pattern Recognition* (CVPR, 2011), pp. 329–336.
39. D. Liu, J. Gu, Y. Hitomi, M. Gupta, T. Mitsunaga, and S. K. Nayar, "Efficient space-time sampling with pixel-wise coded exposure for high-speed imaging," IEEE Trans. Pattern Anal. Machine Intelligence **36**(2), 248–260 (2014).
40. R. Koller, L. Schmid, N. Matsuda, T. Niederberger, L. Spinoulas, O. Cossairt, G. Schuster, and A. K. Katsaggelos, "High spatio-temporal resolution video with compressed sensing," Opt. Express **23**(12), 15992–16007 (2015).
41. 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**, 10526–10545 (2013).
42. J. M. Bioucas-Dias and M. A. Figueiredo, "A new twist: two-step iterative shrinkage/thresholding algorithms for image restoration," IEEE Trans. Image Process. **16**, 2992–3004 (2007).
43. "DLP LightCrafter™ 4500," http://www.ti.com/lsds/ti/dlp/advanced-light-control/microarray-greater-than-1million-lightcrafter4500.page. Accessed: 2016-06-01.
44. C. P. McElhinney, J. B. McDonald, A. Castro, Y. Frauel, B. Javidi, and T. J. Naughton, "Depth-independent segmentation of macroscopic three-dimensional objects encoded in single perspectives of digital holograms," Opt. Lett. **32**, 1229–1231 (2007).

1. Introduction

Recent years have witnessed a great interest in exploiting the redundant nature of signals. The redundancy of acquired signals provides the opportunity to sample data in a compressive approach. Candès et al. [1,2] and Donoho [3] have discussed the high probability of reconstructing

Optics EXPRESS

signals with high fidelity from few random measurements, provided that the signals are sparse or compressible in a known basis. Since then, the theory of compressed sensing (CS) has been widely applied to computational imaging. Lustig et al. [4] described the natural fit of CS to magnetic resonance imaging (MRI). Gan [5] proposed block compressed sensing method for natural images, which is applicable for low-power, low-resolution imaging devices. Brady et al. [6] showed that holography can be viewed as a simple spatial encoder for CS and demonstrated 3D tomography from 2D holographic data.

Gabor's invention of holography in 1948 [7] has provided an effective method for recording and reconstructing a 3D light field from a captured 2D hologram. The use of a CCD camera to digitally record holographic interference patterns has made digital holography (DH) an emerging technology with a variety of imaging applications, such as particle imaging, tracking in biomedical microscopy [8–12] and physical process profiling and measuring [13–17]. Digital Gabor/in-line holography (DIH) is a simple, lensless, yet effective setup for capturing holograms. The simplicity of DIH is balanced by the requirement that objects be small enough to avoid occluding the reference beam significantly [18]. Extensive discussions and applications of DIH have been focused on microscopic imaging, i.e. small and fast-moving objects [19–22]. The tracking of fast movements usually entails multiple exposures [8, 10, 15, 16, 23, 24]. Temporal resolution is usually limited to the 10-100 millisecond range and little research has been conducted on temporal compression. However, in recent years, CS has proved a useful tool to increase the spatial information encoded in DH [6, 25]. Rivenson et al. [26] discussed the application of CS to digital Fresnel holography. Liu et al. [27, 28] and Song et al. [29] improved subpixel accuracy for object localization and enhanced spatial resolution (super-resolution). Furthermore, CS theory has proven successful for recovering scenes under holographic microscopic tomography [30], off-axis frequency-shifting holography [31] as well as millimeter-wave holography [32]. Coded apertures have also been used together with CS to provide robust solutions for snapshot phase retrieval [33, 34]. In view of recent research in CS and DH, several natural questions arise: Can we extend coded aperture to coded exposure? Can we exploit the unused pixels in exchange for increased temporal resolution? Since holography is naturally suitable for recovering depth information, a further research question is whether 4D space-time information can be extracted from 2D data employing the CS framework.

Similar discussions have been initiated in the incoherent imaging regime. Leveraging multiplexing schemes in the temporal domain, e.g. coded exposure, has been demonstrated as an effective hardware strategy for exploiting spatiotemporal trade-offs in modern cameras. High speed sensors usually require high light sensitivity and large bandwidth due to their limited on-board memory. In 2006, Raskar et al. [35] pioneered the concept of coded exposure when he introduced the flutter shutter camera for motion deblurring. The technique requires knowledge of motion magnitude/direction and cannot handle general scenes exhibiting complex motion. Bub et al. [36] designed a high speed imaging system using a DMD (digital micromirror device) for temporal pixel multiplexing. Gupta et al. [37] showed how per-pixel temporal modulation allows flexible post-capture spatiotemporal resolution trade-off. Reddy et al. [38] used sparse representations (spatial) and brightness constancy (temporal) to preserve spatial resolution while achieving higher temporal resolution. Liu et al. [39] used an over-complete dictionary to sparsely represent time-varying scenes. Koller et al. [40] discussed several mask patterns and proposed a translational photomask to encode scene movements extending the work of [41]. These methods have proved successful for reconstructing fast moving scenes by combining cheap low frame-rate cameras with fast spatio-temporal modulating elements. While all of these techniques enable high speed reconstruction of 2D motion, incorporating holographic capture offers the potential to extend the capabilities to 3D motion. Moreover, in many holography setups, the energy from each scene is distributed across the entire detector so that each pixel contains partial information about the entire scene. This offers the potential for improved performance relative to incoherent

Optics EXPRESS

architectures.

Our work exploits both spatial and temporal redundancy in natural scenes and generalizes to a 4D (3D positon with time) system model. We show that by combining digital holography and coded exposure techniques using a CS framework, it is feasible to reconstruct a 4D moving scene from a single 2D hologram. We demonstrate a temporal super resolution of 10×. Note that this increase in frame rate can be achieved for any sensor, regardless of the native frame rate, as long as the spatial-temporal modulator operates at a higher frame rate. We anticipate approximately 1 cm resolution with optical sectioning. As a test case, we focus on macroscopic scenes exhibiting fast motion of small objects (vibrating bars or small particles, etc.).

2. Generalized system model

Digital Gabor holography requires no separation of the reference beam and the object beam. The object is illuminated by a single beam and the portion that is not scattered by the object serves as the reference beam. This concept leads to a simple experimental setup but demands limited object sizes so that the reference beam is not excessively disturbed. In this case, the imaging process is a recording of the diffraction pattern of a 2D aperture.

2.1. Diffraction theory

We first model diffraction in a 2D aperture case. According to Fresnel-Kirchoff diffraction formula [18], the field at each observation point $E(x, y; z)$ in a 2D plane can be written as

$$\begin{aligned} E(x, y; z) &= -\frac{ik}{2\pi} \iint_{\Sigma_0} E_0(x_0, y_0) \frac{\exp(ikr)}{r}\, dx_0\, dy_0, \\ &\approx -\frac{ik}{2\pi z} \iint_{\Sigma_0} E_0(x_0, y_0) \exp\left\{ik\left[(x - x_0)^2 + (y - y_0)^2 + z^2\right]^{\frac{1}{2}}\right\} dx_0\, dy_0, \end{aligned} \tag{1}$$

where r denotes the distance from (x_0, y_0) at the input plane Σ_0,with input field $E_0(x_0, y_0)$, to (x, y) at the output plane, i.e. $r = \left[(x - x_0)^2 + (y - y_0)^2 + z^2\right]^{\frac{1}{2}}$. A further approximation can be made as $r \approx z$ in the denominator based on paraxial approaximation. In the second line of Eq. (1) we make a further approximation of $r \approx z$ in the denominator, but not in the exponent. The integral then becomes a convolution,

$$E(x, y; z) = H * E_0, \tag{2}$$

with the kernel

$$H(x, y; z) = -\frac{ik}{2\pi z} \exp\left[ik\left(x^2 + y^2 + z^2\right)^{\frac{1}{2}}\right]. \tag{3}$$

The kernel H is also referred to as the point spread function (PSF). Since the propagation is along the z-axis, the form of the kernel is determined by the propagation distance z.

2.2. 4D model

We now extend our analysis to a 4-dimensional model. As illustrated in Fig. 1, consider a 4D field $V(x, y, z, t)$, which propagates along the positive z-direction. Along the propagation path, a high-speed coded mask $M(x, y, t)$ is located at z_1. A sensor is placed on the sensing plane z_2. In one frame, the sensor captures the intensity of the field during an exposure time of Δt. The volume can be discretized into N_d planes, with the furthest plane having a distance of d_n with respect to the observation plane at z_0. In Gabor holography, the object beam and the reference beam overlap with each other. This requires the objects to be sparse so that the occlusion of

Optics EXPRESS

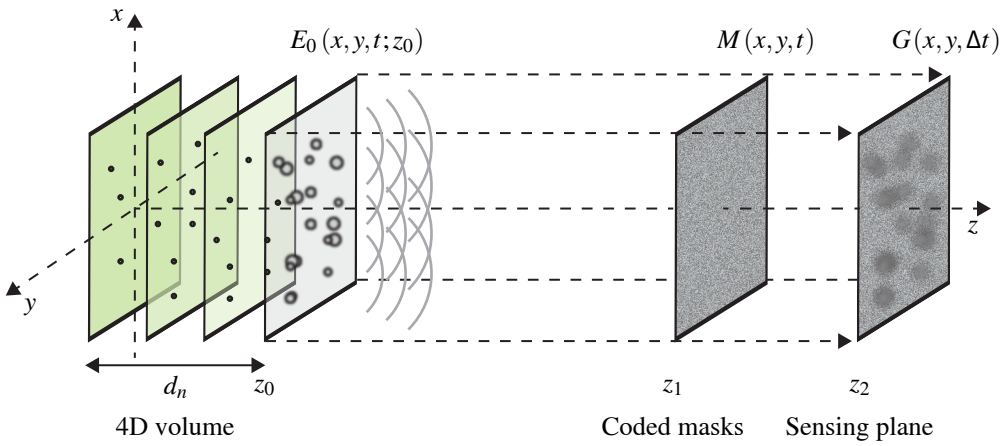

Fig. 1. 4D holographic model. $E_0(x, y, t; z_0)$: projection of a 4D field at z_0, the n-th depth plane has a distance of d_n to z_0; $M(x, y, t)$: temporal coded mask located at z_1; $G(x, y, \Delta t)$: captured image with an integral over Δt. The sensor is located at z_2.

the reference beam is negligible. Under this assumption, the field V in reality represents the summation of the object field O and the constant reference field R. Thus, the field at z_0 is

$$E_0(x, y, t; z_0) = \sum_{n=1}^{N_d} H_{d_n} * O(x, y, z_0 - d_n, t) + R, \tag{4}$$

where H_{d_n} denotes the convolutional kernel for distance d_n.

At the sensing plane, during one exposure time Δt, the sensed image can be expressed as an integral of the intensity I of the field as

$$\begin{aligned} G(x, y, \Delta t) &= \int_{t=t_0}^{t_0+\Delta t} I(x, y, t; z_2) dt \\ &= \int_{t=t_0}^{t_0+\Delta t} \left| H_{z_2-z_1} * \left\{ M(x, y, t) \left[H_{z_1-z_0} * E_0(x, y, t; z_0) \right] \right\} \right|^2 dt, \end{aligned} \tag{5}$$

where $H_{z_1-z_0}$ denotes the propagation process (convolutional kernel) from z_0 to z_1; $H_{z_2-z_1}$ denotes the propagation process (convolutional kernel) from z_1 to z_2; $M(x, y, t)$ denotes a time-variant mask located at z_1. Equation (5) describes the continuous form of the sensing process. However, the mask is operated in a discrete form at high frame rates. Suppose that for each sensor frame the coded mask changes T times at equal intervals of $\tau = \Delta t/T$ during Δt. Then the

Optics EXPRESS

discretized form of G is

$$\begin{aligned} G(x,y,\Delta t) &= \sum_{i=0}^{T-1} \int_{t=t_i}^{t_{i+1}} \left| H_{z_2-z_1} * \left\{ M(x,y,t) \left[H_{z_1-z_0} * E_0(x,y,t;z_0) \right] \right\} \right|^2 dt \\ &= \tau \sum_{i=0}^{T-1} \left| H_{z_2-z_1} * \left\{ M(x,y,t_i) \left[H_{z_1-z_0} * E_0(x,y,t_i;z_0) \right] \right\} \right|^2 \\ &= \tau \sum_{i=0}^{T-1} \left| H_{z_2-z_1} * \left\{ M(x,y,t_i) \left[H_{z_1-z_0} * \left(\sum_{n=1}^{N_d} H_{d_n} * O(x,y,z_0-d_n,t_i) + R \right) \right] \right\} \right|^2 \\ &= \tau \sum_{i=0}^{T-1} \left| O_{c,i} + R_{c,i} \right|^2, \end{aligned} \tag{6}$$

where we denote $O_{c,i}$ and $R_{c,i}$ as the transformed field at the capture plane z_2 for each time frame i.

Then the captured intensity term I can be expanded as $I = |O_c + R_c|^2 = O_c \cdot R_c^* + O_c^* \cdot R_c + O_c^2 + R_c^2$. (Time frame notation i is omitted here.) In [6], Brady et al. neglected the nonlinearity imposed by the squared magnitude and considered the two terms $O_c^2 + R_c^2$ (often referred to as noise and zero-order/DC term) as noise in the measurement model showing that they can be eliminated algorithmically using a CS reconstruction algorithm. In this work, we follow the same approach and the measured intensity can be expressed as

$$I = \{O_c \cdot R_c^* + O_c^* \cdot R_c\} + O_c^2 + R_c^2 = 2Re\{O_c \cdot R_c^*\} + E_E, \tag{7}$$

where E_E combines O_c^2 and R_c^2 into a single term considered as error. We may further assume the reference to be 1 without loss of generality. Then the intensity term can be written as $I = 2Re\{O_c\} + E_E$. In experiment, we approximate the error term by recording the background image and subtract the scene image by this background for reconstruction.

Now we assume that the sensor pixels have the same dimensions as the mask pixels, the unknown field O will have spatial dimensions $N_{M_x} \times N_{M_y}$, depth dimension N_d and temporal dimension T. Further, if we represent the convolutional operations in Eq. (6) as circulant matrices, we can obtain the following compact form

$$\begin{aligned} \mathbf{g} &= 2\mathbf{S}_T Re\left\{ \mathbf{H}_{T,z_{21}} \left\{ \mathbf{M}_T \left[\mathbf{H}_{T,z_{10}} \left(\mathbf{H}_{T,d_n} \mathbf{o} \right) \right] \right\} \right\} + \mathbf{e} + \mathbf{n} \\ &= A(\mathbf{o}) + \mathbf{e} + \mathbf{n}, \end{aligned} \tag{8}$$

where notation and dimensions of the introduced variables are summarized in Table 1 and $A(\cdot)$ describes the complete forward model. Specifically, $\mathbf{S}_T = [I_0, ..., I_{T-1}]$ represents summation over time, where $I_i, i = 0, ..., T-1$ is an identity matrix of size $(N_{M_x} \times N_{M_y}) \times (N_{M_x} \times N_{M_y})$; $M_T = bldg(M_0, M_1, ..., M_{T-1})$ is a block diagonal matrix with each block M_i being a diagonal matrix. Each diagonal element of M_i represents the corresponding pixel operation, e.g. 0 or 1; $\mathbf{H}_{T,z} = [\mathbf{H}_{0,z}, \mathbf{H}_{1,z}, ..., \mathbf{H}_{T-1,z}]$, where z represents z_{21} and z_{10}. $\mathbf{H}_{i,z}$ is a circulant matrix with size $(N_{M_x} \times N_{M_y}) \times (N_{M_x} \times N_{M_y})$ corresponding to the convolutional kernel $H(x,y,z)$. $\mathbf{H}_{T,d_n} = bldg(\mathbf{H}_{0,d_n}, \mathbf{H}_{1,d_n}, ..., \mathbf{H}_{T-1,d_n})$, where $\mathbf{H}_{i,d_n} = [\mathbf{H}_{i,1}, \mathbf{H}_{i,2}, ..., \mathbf{H}_{i,N_d}]$ represents the summation over depths.

In order to reconstruct the 4D volume, an optimization problem is formed as

$$\hat{\mathbf{o}} = \underset{\mathbf{o}}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{g} - A(\mathbf{o})\|_2^2 + \lambda \Phi(\mathbf{o}) \tag{9}$$

Optics EXPRESS

where $\lambda > 0$ is a regularization parameter and $\Phi(\cdot)$ is a regularizer on the unknown 4D field $\mathbf{o}$.

In this work, we employ Total-Variation (TV) as the regularization function defined as

$$\Phi(\mathbf{o}) = \|\mathbf{o}\|_{TV} = \sum_{t=0}^{T-1} \sum_{n=1}^{N_d} \sum_{x=1}^{N_{M_x}} \sum_{y=1}^{N_{M_y}} \left|\nabla (O)_{x,y,n,t}\right|, \tag{10}$$

where we note here that $\mathbf{o}$ is the vectorized version of the unknown 4D object field $O : N_{M_x} \times N_{M_y} \times N_d \times T$. Equation (10) is a generalized 4D TV regularizer. However, the choice of regularizer may vary by different purposes of reconstruction and/or properties of scenes. In experiment, a 3D TV (x, y, n) is used for resolving depths (Section 3.2), i.e., $\Phi_{x,y,n}(\mathbf{o}) = \sum_n \sum_x \sum_y \left|\nabla (O)_{x,y,n}\right|$. TV on temporal domain is included for recovering subtle movement (Section 3.4), i.e., $\Phi_{x,y,t,n}(\mathbf{o}) = \sum_n \sum_t \sum_x \sum_y \left|\nabla (O_n)_{x,y,t}\right|$. Also note that independent regularization parameters may be chosen for the spatial (x, y, n) and time (t) dimensions. We used Two-step IST (TwIST) algorithm [42] for reconstruction.

Table 1. Analysis of all the variables appearing in Eq. (8).

Variable	Description	Dimensions
$\mathbf{g}$	Vectorization of measured intensity G from Eq. (6).	$(N_{M_x} \cdot N_{M_y}) \times 1$
$\mathbf{o}$	Vectorization of unknown 4D object field O.	$(N_{M_x} \cdot N_{M_y} \cdot N_d \cdot T) \times 1$
$\mathbf{e}$	Vectorization of E_E from Eq. (7).	$(N_{M_x} \cdot N_{M_y}) \times 1$
$\mathbf{n}$	Additive measurement noise vector.	$(N_{M_x} \cdot N_{M_y}) \times 1$
$\mathbf{H}_{T,d_n}$	Block diagonal matrix referring to H_{d_n} from Eq. (6). ***Propagation and summation in depth (over N_d) for all time frames (T).***	$(N_{M_x} \cdot N_{M_y} \cdot T)\times$ $(N_{M_x} \cdot N_{M_y} \cdot N_d \cdot T)$
$\mathbf{H}_{T,z_{10}}$	Matrix referring to $H_{z_1-z_0}$ from Eq. (6). ***Propagation of all time frames (T) from z_0 to z_1 (z_{10}).***	$(N_{M_x} \cdot N_{M_y} \cdot T)\times$ $(N_{M_x} \cdot N_{M_y} \cdot T)$
$\mathbf{M}_T$	Block diagonal matrix containing all masks $M(x, y, t_i)$ for all $i = 0 \ldots T-1$ according to Eq. (6). ***Modulation of all time frames (T) with different masks.***	$(N_{M_x} \cdot N_{M_y} \cdot T)\times$ $(N_{M_x} \cdot N_{M_y} \cdot T)$
$\mathbf{H}_{T,z_{21}}$	Matrix referring to $H_{z_2-z_1}$ from Eq. (6). ***Propagation of all time frames (T) from z_1 to z_2 (z_{21}).***	$(N_{M_x} \cdot N_{M_y} \cdot T)\times$ $(N_{M_x} \cdot N_{M_y} \cdot T)$
$\mathbf{S}_T$	Matrix referring to the outer summation of Eq. (6). ***Summation in time (over T).***	$(N_{M_x} \cdot N_{M_y})\times$ $(N_{M_x} \cdot N_{M_y} \cdot T)$

3. Experimental

3.1. Setup

Figure 2 shows the schematic of the experimental setup. The illumination is produced by a diode laser powered by a pulse generator with wavelength of 532 nm. The input beam is expanded and collimated by a neutral density (ND) filter and a collimating lens set (plano-convex lens, $300mm/35mm = 8.57$ magnification, ND filter omitted). In this setup, all the lens are from Thorlabs LSB04. A digital micromirror device (DMD) is used to perform pixel-wise temporal modulation of the light field, similar to [36]. For our experiments, we used the DLP®LightCrafter 4500™from Texas Instruments Inc. The light engine includes a 0.45-inch DMD with > 1million mirrors, each 7.6 μm, arranged in 912 columns by 1140 rows in a diamond pixel array geometry [43]. The DMD is placed approximately 70mm distance away from the objects. An objective lens (single lens with focal length of 125 mm, aperture diameter of 2.54 cm) is placed in front of the CMOS monochromatic sensor and well-aligned with the DMD

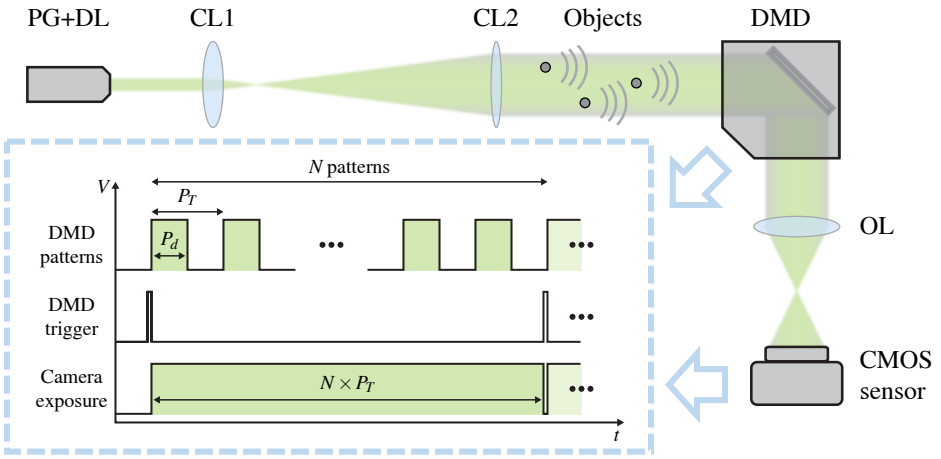

Fig. 2. Schematic of the experimental setup. (PG: Pulse Generator; DL: Diode Laser; CL: Collimating Lens; DMD: Digital Micromirror Device; OL: Objective Lens. A trigger signal generated from the DMD is sent to the camera for exposure. The minimum time between successive DMD mask patterns is $P_T = 500\mu s$ with a pattern exposure $P_d = 250\mu s$. The camera is triggered every N patterns. (N is equal to T in previous context.)

so that it images the DMD plane onto the sensor. The lens introduces a quadratic phase factor inside the integral of Eq. (1). Thus, if the sensor is placed a distance of $2f$ from the OL, the phase is the same as $-2f$ from the lens. In this way, $\mathbf{H}_{T,z_{21}}$ from Eq. (6) reduces to the identity matrix. We used a CMOS monochromatic sensor (Pointgrey GS3-U3-23S6M) with a resolution of 1920×1200 with a pixel pitch of $5.86\mu m$. The key factor is the synchronization between the DMD and the sensor. Each DMD pattern can be projected as fast as $P_T = 500\mu s$ with an effective pattern exposure of $P_d = 250\mu s$. After N patterns are projected, a trigger signal is sent out to the camera which controls the shutter and results in a single exposure.

3.2. *Subsampling holograms*

We start our experiment by examining the reconstruction performance of subsampled holograms. Recovery of a 3D object field from a 2D hologram has been proposed in previous work [6]. The recovery can be treated as inference of high-dimensional data from undersampled measurements. Figure 3 shows the experimental results of 3D recovery with pixel-wise subsampling. For this experiment, we captured two static hairs from craft fur (see below) placed a distance of 7.1 cm and 10.1 cm away from the DMD. Figure 3(a) shows the captured image. To preprocess the captured hologram, first we capture an image on the sensor with no object placed in the field of view - we refer to this as the background image. Note that this captured image corresponds to the term R_c^2 in Eq. (7). We then subtract the hologram by the background image, down-sampled to 960×600 and cropped the central 285×285 ROI around the object. Figure 3(b) shows the captured image of one pattern from the DMD. Each pattern randomly selects 10% of the entire image. To avoid aliasing artifacts caused by the diamond shaped sampling patterns on the DMD, we group together 4×4 adjacent pixels on the DMD to make a single superpixel [43]. Since we are directly imaging on the DMD plane, the resolution is defined by the DMD. In our reconstructions, we design the regularizer to be $\Phi_{x,y,n}(\mathbf{o}) = \sum_n \sum_x \sum_y \left|\nabla (O)_{x,y,n}\right|$, as our focus is the depth. In order to form the matrix A from Eq. (8), we capture images of the mask with no object present. These captured images are divided by the background image to remove the effect of beam non-uniformity. Figure 3(c) shows the subsampled hologram. Figure 3(d) and

Optics EXPRESS

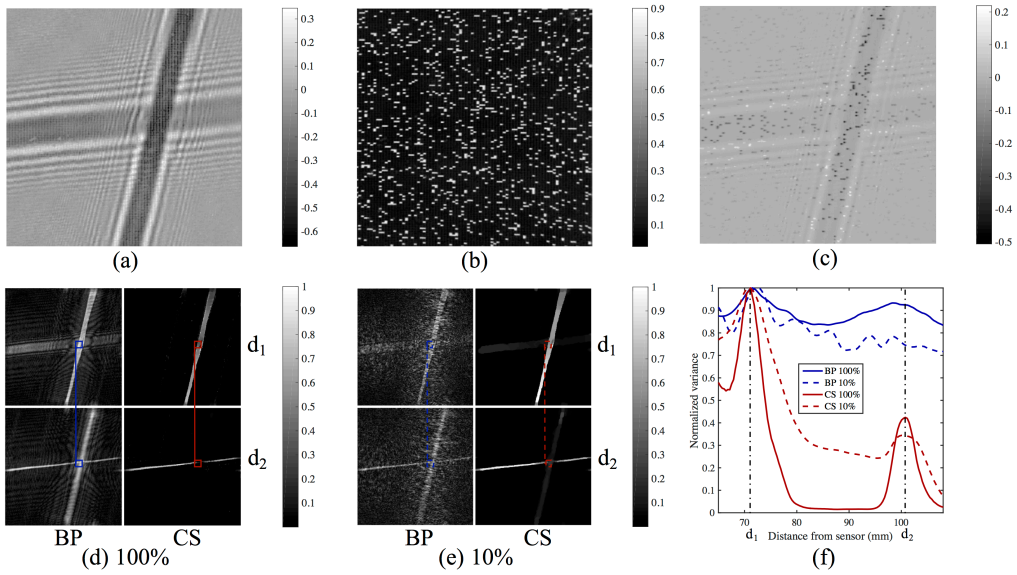

Fig. 3. Subsampling holograms (background subtracted). (a)hologram of two static furs 7.1 cm and 10.1 cm away from sensor. (b) DMD mask, 10%, uniformly random (background divided). (c) subsampled hologram. (d) Comparison of reconstructions from both back-propagation (BP) method and compressed sensing (CS) method using the full hologram. (e) Comparison for BP and CS using 10% subsampled hologram. (f) Normalized variance vs. distance on z direction. Blue series: BP; red series: CS; full curve: 100% hologram; dashed curve: 10% hologram. (See Visualization 1 and Visualization 2.)

Fig. 3(e) compares reconstructions for the full hologram and subsampled hologram. The image (285 × 285) was reconstructed into a 3D volume (285 × 285 × 120) with a depth range from 65 mm to 108 mm. Shown are the images reconstructed at the depth planes corresponding to the location of the two hairs. In order to quantify the performance in terms of depth resolution, we used block variance [44] for the edge pixel of the cross section by the two hairs. Higher variance infers higher contrast, and thus, higher resolution. The block variance was computed within a window of 21 × 21 pixels highlighted as blue and red in Figs. 3(d) and 3(e). Figure 3(f) shows the normalized variance versus depth from sensor. Two principle peaks are observed and can be inferred as the focus distance for the two furs. The peak around d_1 has strong signal in all four curves. This was because the object located there has larger size than the other one. As can be seen, using only 10% of the data deteriorates both BP and CS reconstruction resolutions. And in 10% from BP, it is even harder to track the second object because of the impact of mask pattern. This can also be observed in the left panel of Fig. 3(e) where the back propagation of the mask severely affected the objects. The variance decreases fast in CS reconstructions. This implies the denoising effect as well as the optical sectioning power of CS. In 10% reconstruction, the intermediate volume between the two objects were not denoised as good as in "100%" case. This shows that greater subsampling factors reduce the effective depth resolution.

3.3. *Temporal multiplexing*

In the previous section, we analyzed the effect of subsampling on reconstruction performance for compressive holography. Here we show how to utilize the excess pixel bandwidth in the sensor to increase temporal resolution. A simulation experiment was carried out in order to

Optics EXPRESS

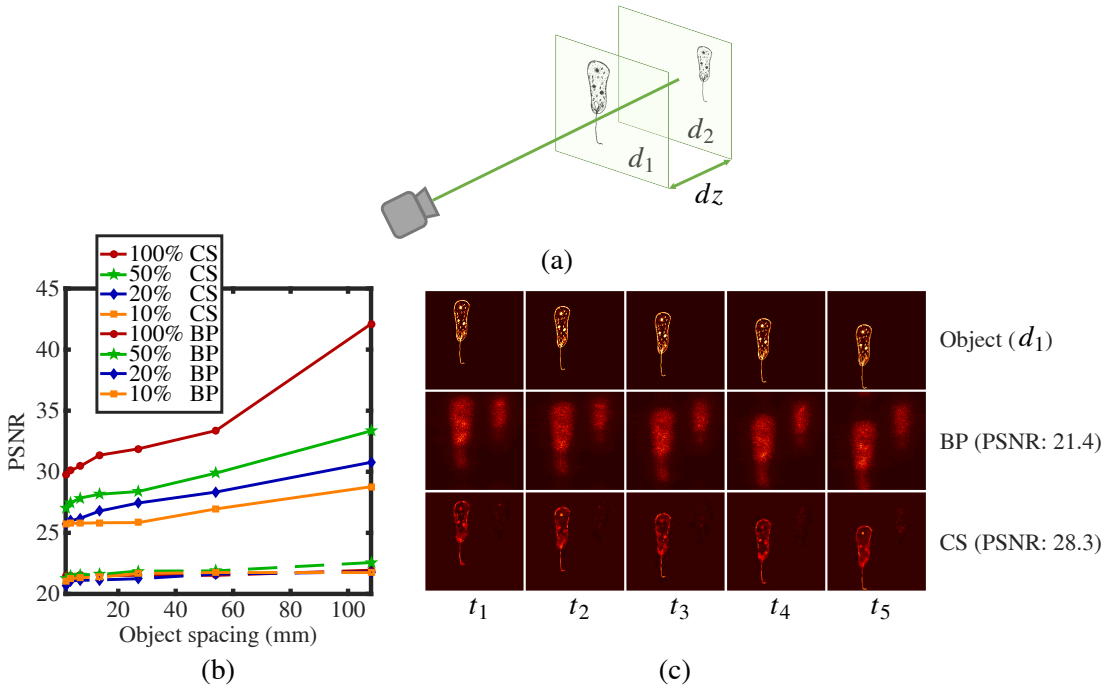

Fig. 4. Performance simulations. (a) Scenario: two Peranema with different sizes moving at different planes (dz), a single image is simulated at the sensor plane; (b) Space-time performance. Horizontal axis indicates different spacing between the two objects. "100%": full resolution; "50%": 50% of the pixels are randomly sampled at each time frame, which corresponds to a temporal increase of 2; "20%": temporal increase of 5; "10%": temporal increase of 10. Lines represent CS results and dashed lines represent BP results. PSNR in dB. (c) Reconstruction results at depth d_1. Marked as red circle in (b).

quantitatively analyze our imaging system (Fig. 4). As shown in Fig. 4(a), two layers of objects (peranema with different scales) were used as a test case. Each layer had 256×256 pixels. The pixel pitch was set so that the whole scene size ($9.85 \times 9.85mm$) was approximately identical to the DMD size. The first object was placed at $70mm$ away from the sensor. The other object was placed dz further away from the first object. dz is a changing variable. A spatiotemporal subsampling mask was displayed on the DMD. For example, when n time frames are required, each frame will have $1/n \times 100\%$ of the pixels being randomly selected and displayed. In this way, the summation of n frames is the full resolution scene image. In simulation, we omitted the propagation between the DMD and the sensor. For reconstruction, we compared back-propagation and compressed sensing. In order to have a better reconstruction result, we inserted 4 intermediate planes between the two objects. The results are shown in Figs. 4(b) and 4(c). In Fig. 4(b), the peak signal-to-noise ratio (PSNR) was used to measure the reconstruction performance. $PSNR = 10log_{10}(peakvalue^2/MSE)$, where MSE is the mean-squared error between the reconstruction and the input object field. The PSNR is computed on the 4D volume, which can also be treated as an average over multiple time frames. The higher PSNR value is, the better fidelity the reconstruction is. We picked out a point from Fig. 4(b), marked as red ring, to show in Fig. 4(c) the visual meaning of the PSNR values. It can be seen that lower rate of subsampling causes worse reconstruction performance. PSNR also decreases with the decrease of object spacing.

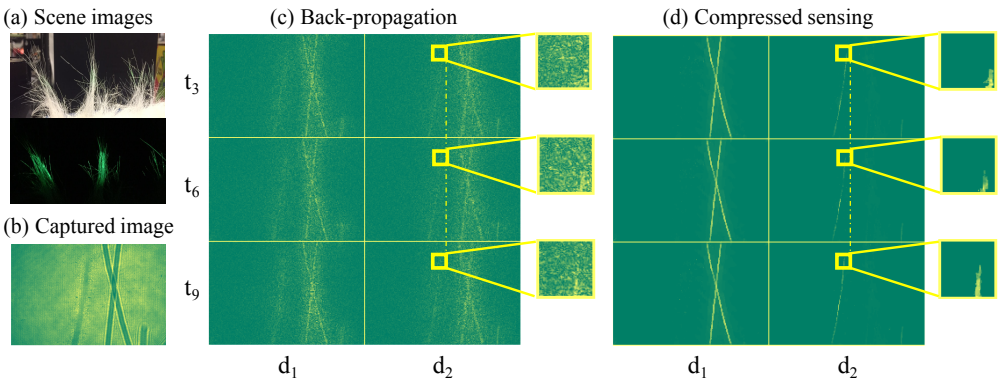

Fig. 5. Reconstruction results from a single image (moving hairs). 10 frames of video and two depth frames are reconstructed from a single captured hologram. Due to space constraints, 3 video frames (3rd, 6th, 9th) and two depths ($d_1 = 73mm, d_2 = 111mm$) are presented (see Visualization 3 and Visualization 4).

3.4. *Spatiotemporal recovery for fast-moving objects*

We present two illustrative examples which are aimed for the application of observing subtle vibrations and tracking small-but-fast-moving particles.

Figure 5 shows a reconstruction result demonstrating a 10× increase in temporal resolution. The captured image contains several strands of hair blown by an air conditioner. *From a single captured image, we reconstruct 2 depth slices and 10 frames of video*. In the case of small lateral movement, i.e. vibration, it is feasible to apply total variation on time domain, i.e., $\Phi_{x,y,t,n}(\mathbf{o}) = \sum_n \sum_t \sum_x \sum_y \left|\nabla (O_n)_{x,y,t}\right|$. In this case, the depths of the objects are pre-determined. For the convenience of comparison, 3 time frames (3rd, 6th, 9th) are shown for both back-propagation and compressed sensing. In terms of depth, our CS result shows well-separated objects at different depth layers while the back-propagation method fails to achieve optical sectioning. The movement of the object is also recovered in our CS result. The reconstruction was performed on a computer with Intel Core i5 CPU at 3.2 GHz and 24 GB of RAM. The data processing takes about 2.4 hours for A with the size of $(960 \times 600) \times (960 \times 600 \times 2 \times 10)$. The codes were written in Matlab 2015b.

Figure 6 shows another reconstruction result for dropping several flakes of glitter. The glitter flakes in Fig. 6(a) had size of 1 mm and were dropped in a range of 60 mm to 80 mm away from sensor. The glitter flakes were also blown by an air conditioner. Figure 6(b) shows the captured single image. Figure 6(c) shows preprocessed image which is subtracted by background image. In this case, the glitter flakes were moving at high speed. There was no overlap between two consecutive frames for the same flake. So each frame was recovered independently. For each frame, a depth range was estimated with 120 layers. The regularizer is designed to be $\Phi_{x,y,n}(\mathbf{o}) = \sum_n \sum_x \sum_y \left|\nabla (O)_{x,y,n}\right|$. Figure 6(d) shows a reconstruction map of 2 depths and 4 time frames. The downward and leftward motion of two glitter flakes can be observed. A similar refocusing method was used as in [44]. Here, we scanned the reconstructed image by a 21×21 window and computed the variance (normalized) to get the focused depth information. If the normalized variance at defocused depth are higher than 0.5, that pixel was rejected as background/noise. For adjacent pixels which have similar variance profile, the pixels were treated as a single particle. Figure 6(e) shows the normalized variance for two particles at d_1 and d_2. The particles are tracked at two locations pointed out by the arrows in Fig. 6(d). The

Optics EXPRESS

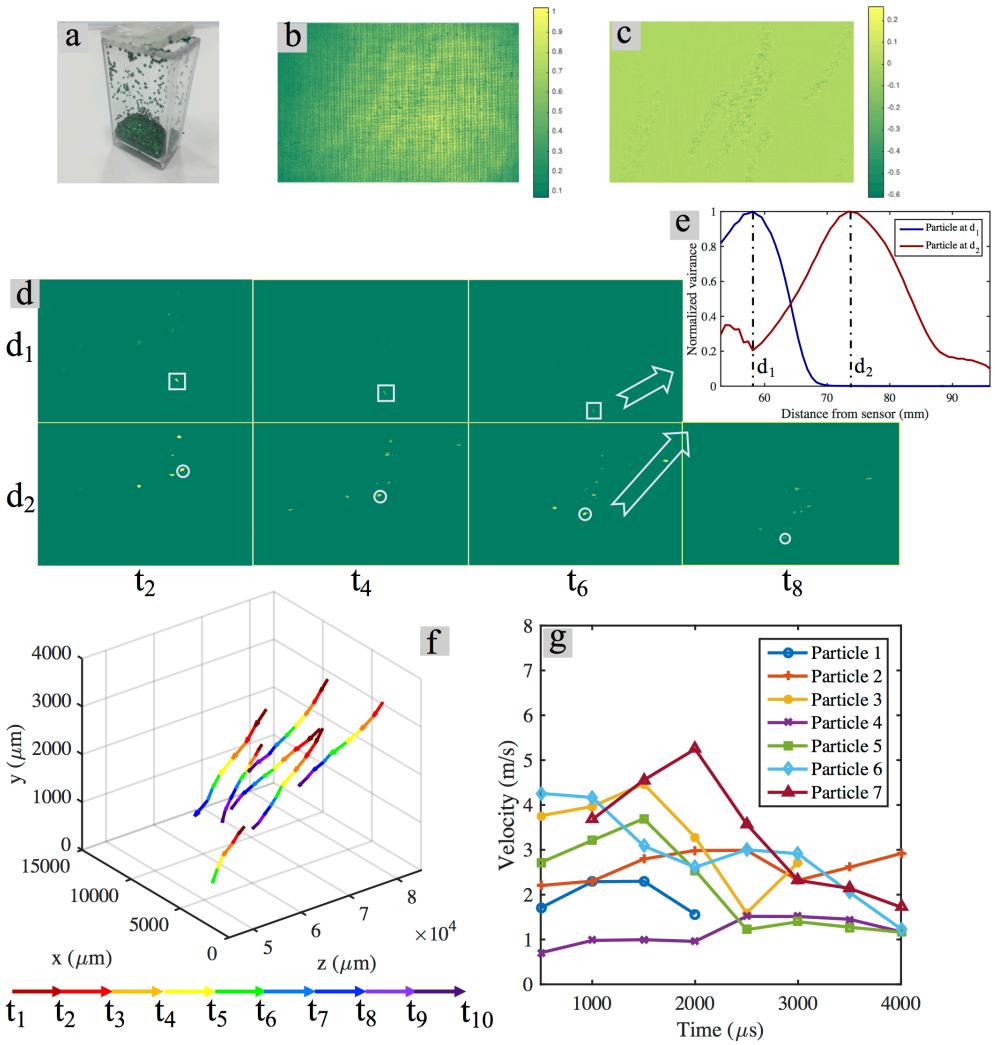

Fig. 6. Reconstruction results from a single image (dropping glitters). (a) Glitters; (b) captured image; (c) normalized image; (d) reconstruction map. 2 depths and 4 out of 10 frames are shown; (e) normalized variance plot from 2 particles at d_1 and d_2; (f) 4D particle tracking; (g) velocity plotting with time range from $500\mu s$ to $4000\mu s$.

overall tracking results are shown in Fig. 6(f) and 6(g). 7 particles are detected with 4D motion within 5 ms. In Fig. 6(f), the temporal transition was represented by 10 different color arrows. Figure 6(g) shows a velocity chart of the 7 particles. The velocity of each particle was computed by $v(t_n) = [d(t_{n+1}) - d(t_{n-1})]/2\Delta t$, where $d(t_n)$ depicts the 3D location at n-th time frame, $\Delta t = 500\mu s$. The velocity of the particles ranges from 0.7 m/s to 5.5 m/s. In this case, each time frame was processed separately. The data processing time for each frame takes about 7 hours for A with the size of $(960 \times 600) \times (960 \times 600 \times 60)$.

Optics EXPRESS

4. Discussion

We have demonstrated two illustrative cases where 4D spatio-temporal data is recovered from a single 2D data. In the case of vibrating hairs, 2 depth layers and 10 video frames in time were recovered. The spatio-temporal compression is 20×. In the case of dropping glitter flakes, a 4D volume was reconstructed to track the motion of small particles. The spatio-temporal compression is 120 × 10. We call our technique "compressive holographic video" to emphasize the compressive sampling approach to acquisition of spatio-temporal information. We show that our technique affords a significant reduction in space-time sampling, enabling 4D events to be acquired using only a single captured image.

In our prototype implementation we use a DMD as a coded aperture that is imaged directly onto a sensor. While non-trivial to implement, in principle it is possible to fabricate a CMOS sensor with pixel-wise coded exposure control. The prototype showed that it is possible to simultaneously exceed the capture rate of imagers and recover multiple depths with reasonable depth resolution. In this paper, as an example, we presented a temporal increase factor of 10×. A potential factor can be 24× based on the DMD we used. By means of spatio-temporal modulator, one is able to significantly increase the frame rate of the sensors. Based on this idea, the recovered frame rate is redefined by the modulator's frame rate. The coded-exposure technique enables high speed imaging with a simple frame rate camera. Digital in-line holography brings the capability of 3D tomographic imaging with simple experimental setup. Our Compressive Holographic Video technique is also closely related to phase retrieval problems commonly faced in holographic microscopy. Our space-time subsampling technique can be viewed as a sequence of coded apertures applied to a spatiotemporally varying optical field. In the future we plan to explore the connections between our CS reconstruction approach and the methods introduced in [33]. In our general model, we place a coded aperture between the sensor and scene. In our prototype implementation we use a DMD as a coded aperture that is imaged directly onto a sensor. While not explored in this paper, we believe that adding defocus between the coded aperture plane and sensor may be beneficial for phase retrieval tasks, as in [33]. In this work, we focus on a proof-of-principle demonstration of compressive holographic video. In the future, we hope to explore a diverse set of mask designs, as well as techniques for mask optimization.

Funding

National Science Foundation (NSF) CAREER grant IIS-1453192;
Office of Naval Research (ONR) grant 1(GG010550)//N00014-14-1-0741;
Office of Naval Research (ONR) grant #N00014-15-1-2735.

Acknowledgments

The authors were grateful for the constructive discussions with Dr. Roarke Horstmeyer, Donghun Ryu and the reviewers.

185.124.145.29 {ts '2017-01-05 18:48:05'}