Dictionary-based phase retrieval for space-time super resolution using lens-free on-chip holographic video

Zihao Wang[1,*], Qiqin Dai[1], Donghun Ryu[2], Kuan He[1], Roarke Horstmeyer[3], Aggelos K. Katsaggelos[1], Oliver Cossairt[1]

[1] *Electrical Engineering and Computer Science, Northwestern University, Evanston, IL 60201, USA*
[2] *Electrical Engineering, California Institute of Technology, Pasadena, CA 91125, USA*
[3] *Charité Medical School, Humboldt University of Berlin, Berlin 10117, Germany*
[*] *Corresponding author: zwinswang@gmail.com*

Abstract: We propose a dictionary-based phase retrieval approach for monitoring in vivo biological samples based on lens-free on-chip holographic video. Our results present a temporal increase of 9× with 4×4 sub-sampling.

OCIS codes: 090.1995, 100.5070, 110.1758.

1. Introduction

On-chip in-line holography (OIH) is a lens free imaging method that offers high resolution over a wide field-of-view [1, 2]. However, high spatial resolution requires longer time for sampling and processing, yielding a low frame rate. As the frame rate increases, spatial resolution decreases. This limitation is mainly caused by hardware factors, such as the readout and Analog-to-Digital (A/D) conversion time of sensors. A typical strategy for increasing the frame rate is implemented by pixel-wise sub-sampling [3, 4]. In our case (IDS UI-148xLE CMOS sensor with Intel core i5, 4GB RAM), the sensor is capable of performing periodic sub-sampling (static over time) with different factors. The frame rate of the sensor can be increased up to 14× (61.5 FPS) from the full resolution (1920×2560 pixels) with a down-sampling factor of 36 (6×6 periodic). However, the high frequency information of sub-sampled images will be severely affected, resulting in an inaccurate reconstruction.

In this work, we aim at overcoming the space-time resolution tradeoff using a dictionary-based single frame super resolution approach. Our result demonstrates a suitable performance for space-time reconstruction with a temporal increase of 9× with 4×4 sub-sampling. Further, we compare phase retrieval performance with previously proposed method, SPR [2]. This dictionary-based super resolution scheme is applicable for monitoring *in vivo* movement of biological samples, such as Euglena.

2. Methods

A simple on-chip imaging setup is shown in Fig. 1(a). The system consists of a light-emitting diode (LED) and a complementary metal-oxide-semiconductor (CMOS) sensor. The LED (quasi-monochromatic with wavelength $\lambda = 625\mu m$) is positioned at approximately 40 cm above the sensor so that the spatial coherence of the source, defined by its distance from the sensor and the width of its active area, is sufficient to produce high-contrast diffraction fringes at the sensor plane. The samples for imaging are prepared on a transparent glass. The glass is closely placed and adjusted parallel to the sensor plane. The field propagation process is modeled as,

$$H(x,y;z) = \mathfrak{F}^{-1}\left\{\mathfrak{F}\{E(x,y)\}[k_x,k_y]\exp\left[iz\sqrt{k^2-k_x^2-k_y^2}\right]\right\}[x,y], \tag{1}$$

where $H(x,y;z)$ is the propagated field at axial distance z from the object field $E(x,y)$. $\mathfrak{F}$ and $\mathfrak{F}^{-1}$ denote the Fourier transform and its inverse transform. k_x and k_y are the corresponding spatial frequencies of x and y. $k = 2\pi/\lambda$. Previous reconstruction techniques [5, 6] alternatively project the image field between Fourier domain ($H(x,y)$) and object domain ($E(x,y)$). The corresponding constraints, i.e., amplitude in Fourier domain and support in object domain, are updated accordingly. Ryu et al. [2] reported the issue of space-time resolution tradeoff and proposed a sub-sampled phase retrieval (SPR) scheme for temporal resolution enhancement. Here, we propose a dictionary-based method for further improving the reconstruction results. The reconstruction is decoupled into two steps. The first step is HR hologram recovery via sparse representation, as shown in Fig. 1(c). A bilevel dictionary is trained from acquired high spatial resolution (HR) holograms and their sub-sampled low spatial resolution (LR) counterparts. The second step is to retrieve phase based on the recovered HR holograms.

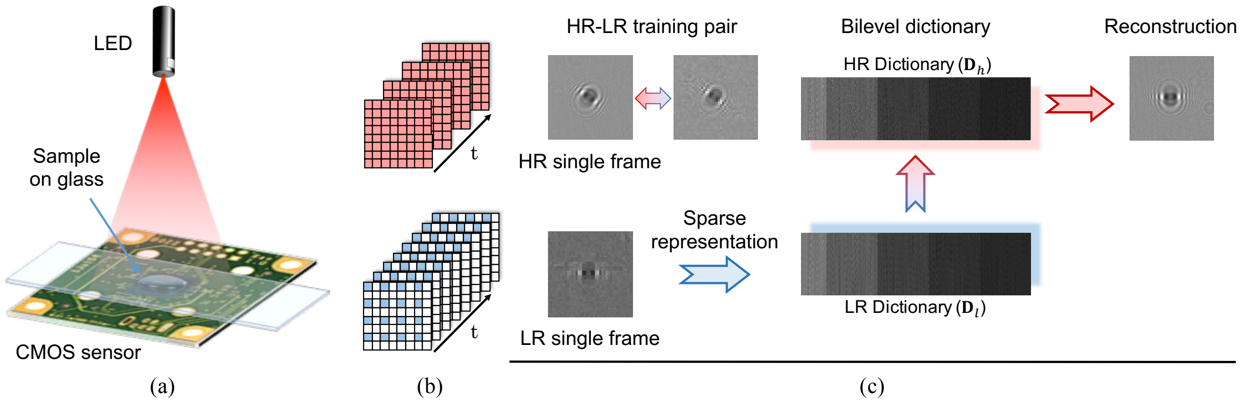

Fig. 1. (a) Experimental setup. LED: light-emitting diode; CMOS: complementary metal-oxide-semiconductor. (b) Space-time sampling. Up: full/high spatial resolution with low temporal resolution; down: sub-sampled/low spatial resolution with improved temporal resolution. (c) Bilevel dictionary-based training/reconstruction procedure. HR: high resolution; LR: low resolution.

The problem in sparse signal representation is to find the sparsest representation possible of a given signal vector $y \in \mathbb{R}^n$, based on an over-complete dictionary $\Phi \in \mathbb{R}^{n\times m}$, with $m > n$. Each column vector of the dictionary $\phi_i \in \mathbb{R}^n$, $i = 1,...,m$. is referred to as an atom. Thus, the sparse representation problem becomes the following optimization problem, $\min_\alpha ||\alpha||_0 \quad s.t. \quad \mathbf{y} = \Phi\alpha$. This is a combinatorial optimization problem. Several variants [7, 8] such as convex relaxation have been proposed. In order to avoid the under-optimized connection between LR image and HR dictionary, we use a sparse coding scheme, bilevel coupled dictionary [9, 10] to jointly optimize an HR dictionary and its corresponding LR dictionary, as well as the sparse representation coefficients. The optimization problem can be described as below,

$$\begin{aligned} &\min_{\mathbf{D}_l,\mathbf{D}_h} \sum_{i=1}^{N} \frac{1}{2}||\mathbf{H}_i - \mathbf{D}_h\alpha_i^H||_2^2 \\ s.t.\ \alpha_i^H &= \arg\min_{\alpha_i^L} \frac{1}{2}||\mathbf{L}_i - \mathbf{D}_l\alpha_i^L||_2^2 + \lambda||\alpha_i^L||_1 \end{aligned} \tag{2}$$

Each element of the HR and LR dictionaries obeys the constraint: $||\mathbf{D}_h(:,k)||_2 \leq 1, ||\mathbf{D}_l(:,k)||_2 \leq 1, k = 1,...,m$. The two dictionaries are coupled so as to share the same sparse coefficient $\alpha_i^H = \alpha_i^L$.

We build the coupled dictionary separately for each sub-sampling factor. In the training phase, we record a training video in order to acquire HR image patches (10^6). The corresponding LR image patches are obtained by numerically sub-sampling. The number for dictionary atom is 512. In the testing phase, the same 800 image patches were used. Empirically, we found that larger image patch sizes improve the reconstruction performance. However, larger image patch sizes require more atoms and larger number of training patches. The training results are summarized in Table 1.

3. Results and conclusion

A comparison of reconstruction performances is shown in Fig. 2. In this case, the parameters, e.g., distance of propagation, are fixed. The parameters in the phase retrieval algorithm are approximately the same after tuning. In order to perform a fair comparison, the central 60×60 pixels are cropped from the originally reconstructed 160×160 image and the PSNR and SSIM values are computed based on the cropped images. It can be seen that the dictionary-based method performs better than SPR algorithm.

	2×2	3×3	4×4
5×5 pixels	38.78	N.A.	$32.10^{\dagger}$
7×7 pixels	39.04	35.67	32.33
9×9 pixels	39.13	36.72	33.13

Table 1. Comparison of different patch sizes at different sub-sampling factors. PSNR values are shown in decibel unit. †: tested by applying the 2×2 dictionaries twice.

Our proposed method serves as a useful tool for improving space-time resolution based on OIH setup. Monitoring *in vivo* scenes also provides insights for behavior studies of microorganism.

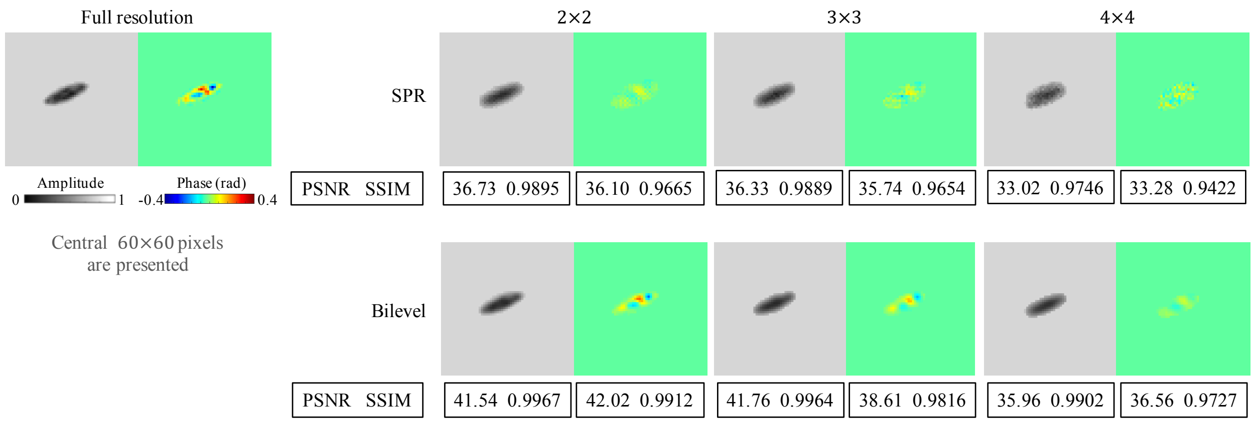

Fig. 2. Comparison of sub-sampled phase retrieval (SPR) [2] and our proposed bilevel dictionary method. PSNR: peak signal-to-noise ratio; SSIM: structural similarity.

References

1. T. Su, L. Xue, and A. Ozcan, ”High-throughput lensfree 3D tracking of human sperms reveals rare statistics of helical trajectories,” Proceedings of the National Academy of Sciences 109, no. 40 16018-16022 (2012).
2. D. Ryu, Z. Wang, K. He, R. Horstmeyer, and O. Cossairt, ”Subsampled Phase Retrieval for On-chip Lensless Holographic Video,” arXiv preprint arXiv:1612.02122 (2016).
3. 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 transactions on pattern analysis and machine intelligence, 36, 248-260 (2014).
4. Z. Wang, L. Spinoulas, K. He, L. Tian, O. Cossairt, A. K. Katsaggelos, and H. Chen, ”Compressive holographic video,” Opt. Express 25, 250-262 (2017).
5. R. Gerchberg and W. O. Saxton, ”Phase determination from image and diffraction plane pictures in electron-microscope,” Optik 34, no. 3 275 (1971).
6. J. R. Fienup, ”Reconstruction of an object from the modulus of its Fourier transform,” Opt. Lett. 3, 27-29 (1978).
7. D. L. Donoho, M. Elad, and V. N. Temlyakov, ”Stable recovery of sparse overcomplete representations in the presence of noise,” IEEE Transactions on information theory, 52, 6-18 (2006).
8. M. Aharon, M. Elad, and A. Bruckstein, ”K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation.” IEEE Transactions on signal processing, 54, 4311-4322 (2006).
9. J. Yang, J. Wright, T. S. Huang, and Y. Ma, ”Image super-resolution via sparse representation,” IEEE transactions on image processing, 19, 2861-2873 (2010).
10. J. Yang, Z. Wang, Z. Lin, X. Shu, and T. Huang, ”Bilevel sparse coding for coupled feature spaces,” In Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, pp. 2360-2367 (2012).