

Ultra-high resolution coded wavefront sensor

CONGLI WANG, XIONG DUN, QIANG FU, AND
WOLFGANG HEIDRICH*

Visual Computing Center, King Abdullah University of Science and Technology (KAUST), Thuwal
23955-6900, Saudi Arabia
*wolfgang.heidrich@kaust.edu.sa

Abstract: Wavefront sensors and, more general phase retrieval methods have recently attracted a lot of attention in a host of application domains, ranging from astronomy to scientific imaging and microscopy. In this paper, we introduce a new class of sensor, the *Coded Wavefront Sensor*, which provides high spatio-temporal resolution using a simple masked sensor, under white light illumination. Specifically, we demonstrate megapixel spatial resolution and phase accuracy better than 0.1 wavelengths at reconstruction rates of 50 Hz or more, thus opening up many new applications from high-resolution adaptive optics to real-time phase retrieval in microscopy.

© 2017 Optical Society of America

OCIS codes: (010.7350) Wave-front sensing; (100.3190) Inverse problems; (100.5070) Phase retrieval; (110.1758) Computational imaging; (150.4620) Optical flow.

References and links

1. B. C. Platt and R. Shack, "History and principles of Shack-Hartmann wavefront sensing," *Journal of Refractive Surgery* **17**, S573–S577 (2001).
2. L. K. Saddlemyer, G. Herriot, J.-P. Véran, and J. M. Fletcher, "Design aspects of the reconstructor for the Gemini adaptive optics system (Altair)," *Proc. SPIE* **3353**, 150–159 (1998).
3. R. Ragazzoni, "Pupil plane wavefront sensing with an oscillating prism," *J. Mod. Opt.* **43**, 289–293 (1996).
4. O. Fauvarque, B. Neichel, T. Fusco, S. Thierry, and J.-F. Sauvage, "Variation around a pyramid theme: optical recombination and optimal use of photons," *Opt. Lett.* **40**, 3528–3531 (2015).
5. A. F. Brooks, T.-L. Kelly, P. J. Veitch, and J. Munch, "Ultra-sensitive wavefront measurement using a Hartmann sensor," *Opt. Express* **15**, 10370–10375 (2007).
6. F. Roddier, "Curvature sensing and compensation: a new concept in adaptive optics," *Appl. Opt.* **27**, 1223–1225 (1988).
7. L. Waller, L. Tian, and G. Barbastathis, "Transport of intensity phase-amplitude imaging with higher order intensity derivatives," *Opt. Express* **18**, 12552–12561 (2010).
8. Z. Jingshan, R. A. Claus, J. Dauwels, L. Tian, and L. Waller, "Transport of intensity phase imaging by intensity spectrum fitting of exponentially spaced defocus planes," *Opt. Express* **22**, 10661–10674 (2014).
9. M. R. Teague, "Deterministic phase retrieval: a green's function solution," *J. Opt. Soc. Am.* **73**, 1434–1441 (1983).
10. T. E. Gureyev, A. Pogany, D. M. Paganin, and S.W. Wilkins, "Linear algorithms for phase retrieval in the Fresnel region," *Opt. Commun.* **231**, 53–70 (2004).
11. R. Tumbar, R. A. Stack, D. J. Brady, "Wave-front sensing with a sampling field sensor," *Appl. Opt.* **39**, 72–84 (2000).
12. J. C. Chanteloup, "Multiple-wave lateral shearing interferometry for wave-front sensing," *Appl. Opt.* **44**, 1559–1571 (2005).
13. B. K. Horn and B. G. Schunck, "Determining optical flow," *Artificial intelligence* **17**, 185–203 (1981).
14. T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, "High accuracy optical flow estimation based on a theory for warping," in *ECCV'04, 8th European Conference on Computer Vision* (2004) pp. 25–36.
15. S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning* **3**, 1–122 (2011).
16. J. W. Goodman, *Introduction to Fourier Optics* (Roberts and Company Publishers, 2005).
17. K. Matsushima and T. Shimobaba, "Band-limited angular spectrum method for numerical simulation of free-space propagation in far and near fields," *Opt. Express* **17**, 19662–19673 (2009).
18. R.G. Lane, A. Glindeemann, and J.C. Dainty, "Simulation of a Kolmogorov phase screen," *Waves in random media* **2**, 3, 209–224 (1992).
19. R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," *Computer Science Technical Report CSTR* **2**, 1–11 (2005).
20. A. Veeraraghavan, R. Raskar, A. Agrawal, A. Mohan, and J. Tumblin, "Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing," *ACM Trans. Graph.* **26**, 69 (2007).

-
21. K. Marwah, G. Wetzstein, Y. Bando, and R. Raskar, “Compressive light field photography using overcomplete dictionaries and optimized projections,” *ACM Trans. Graph.* **32**, 46 (2013).
22. H. Richard and M. Raffel, “Principle and applications of the background oriented schlieren (bos) method,” *Measurement Science and Technology* **12**, 1576 (2001).
23. B. Atcheson, W. Heidrich, and I. Ihrke, “An evaluation of optical flow algorithms for background oriented schlieren imaging,” *Experiments in Fluids* **46**, 467–476 (2009).
-

1. Introduction

Wavefront sensing and phase retrieval are two closely related tasks that find applications in a wide range of fields. The classical Shack-Hartmann Wavefront Sensor [1] is commonly used in astronomy and ophthalmology. As a first-order method that tracks the 2D motion of focus spots generated by a lenslet array, the Shack-Hartmann sensor offers high frame rates, but suffers from low spatial resolution, corresponding to the number of sub-apertures rather than the number of pixels in the image sensor. While the spatial resolution can be improved by reducing the number of pixels per sub-aperture, e.g. to 2×2 pixels per lenslet as in the Altair adaptive optics system for Gemini [2], this comes at the cost of reduced range, such that large distortions can not be measured. Similar tradeoffs exist for other designs, including the pyramid wavefront sensor [3] and its variant [4], and the Hartmann wavefront sensor [5].

Better accuracy and higher spatial resolution can be achieved with Curvature Wavefront Sensors [6], and closely related phase retrieval methods developed recently in microscopy [7, 8]. These methods are based on inverting the Transport of Intensity Equation (TIE) [9], which requires phase diversity, i.e. multiple image measurements along the optical axis. TIE has been extensively investigated [10], and proves to be effective and accurate approach for wavefront sensing. However, the requirement for phase diversity makes it difficult to design snapshot-capable systems that can image fast-moving dynamic phenomena.

In this paper, we propose a novel first-order wavefront sensor named Coded Wavefront Sensor. By placing a binary mask in close proximity of a camera sensor, one can numerically decode the wavefront from the apparent motion of the diffraction patterns. Our work is closely related to the sampling field sensor [11] and the quadri-wave lateral shearing interferometer wavefront sensor [12], but offers full sensor resolution under both coherent and incoherent illumination. Compared to the classical Shack-Hartmann wavefront sensor, the Coded Wavefront Sensor offers much higher pixel utilization and full sensor resolution reconstruction, overcoming the trade-off between spatial resolution and range. Specifically, the Coded Wavefront Sensor offers the ability to measure large distortions with high spatial resolution. Compared to TIE-based methods, the Coded Wavefront Sensor does not require phase diversity, works with both monochromatic and polychromatic illumination, is simple to build, and offers much faster reconstruction than video-rates. It is therefore highly suitable for imaging dynamic deformations of the wavefront.

2. Principle

A schematic of Coded Wavefront Sensor is shown in Fig. 1. A uniform random binary mask is placed closely (for instance, distance $z \approx 1.5$ mm) in front of a bare image sensor. For calibration, the Coded Wavefront Sensor is illuminated by a planar wavefront, and the corresponding diffraction pattern of the mask on the image sensor is recorded. During measurement, distortions of the wavefront result in localized diffraction pattern displacements (small arrows in Fig. 1(b)), which can be tracked and used to measure the wavefront at the full resolution of the image sensor.

To model the principle of the Coded Wavefront Sensor, we denote the scalar field at the mask plane and the sensor plane as $u_0(\mathbf{r})$ and $u_z(\mathbf{r})$ respectively, where $\mathbf{r} = (x, y)$ denotes a coordinate point on the 2D plane. The original scalar field $u_0(\mathbf{r})$ is the product of the mask transfer function $p_0(\mathbf{r})$ and the scalar optical field $\exp[j\phi(\mathbf{r})]$, where $\phi(\mathbf{r})$ is the distorted wavefront under investigation.

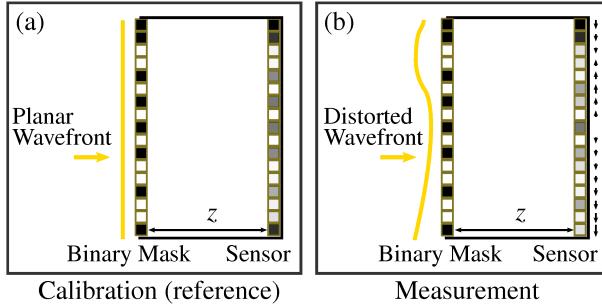


Fig. 1. Schematic of the Coded Wavefront Sensor. (a) A simple calibration is performed by capturing the diffraction pattern of a planar wavefront, namely the reference image. (b) By just recording the diffraction pattern of a distorted wavefront (i.e. the measurement image) and compare it to the reference image captured in (a), the distorted wavefront can be reconstructed. Small arrows indicate local distortion directions.

For mask functions $p_0(\mathbf{r})$ with small feature sizes, the effect of diffraction must be considered, whereas for low frequency wavefront $\phi(\mathbf{r})$, one may simply employ ray optics. Under proper approximation conditions, in the Appendix we show:

$$u_z(\mathbf{r}) \approx \exp[j\phi(\mathbf{r})]p_z(\mathbf{r} - (z/k)\nabla\phi(\mathbf{r})), \quad (1)$$

where $k = 2\pi/\lambda$ is the wave number for wavelength λ , and $p_z(\mathbf{r})$ is the diffraction field of $p_0(\mathbf{r})$ under collimated illumination.

Denoting the calibration (reference) image and the measurement image as $I_0(\mathbf{r})$ and $I(\mathbf{r})$, given $I_0(\mathbf{r}) = |p_z(\mathbf{r})|^2$, the measurement image $I(\mathbf{r})$ is shifted relative to $I_0(\mathbf{r})$ by a point-wise apparent motion proportional to the wavefront slope $\nabla\phi(\mathbf{r})$:

$$I(\mathbf{r}) = I_0(\mathbf{r} - (z/k)\nabla\phi(\mathbf{r})). \quad (2)$$

Eq. (2) highlights the underlying principle of our Coded Wavefront Sensor: the distorted wavefront results in apparent motion of the diffraction pattern, assuming no scintillation. Note that the apparent motion is irrelevant to the wave number k if we consider the distorted wavefront $\phi(\mathbf{r}) = ko(\mathbf{r})$ with optical path $o(\mathbf{r})$, which means the Coded Wavefront Sensor allows for broadband illumination.

However, Eq. (2) is nonlinear. To retrieve $\phi(\mathbf{r})$ from $I_0(\mathbf{r})$ and $I(\mathbf{r})$, one may iteratively solve a linearized version of Eq. (2). At each step, the linearized version of Eq. (2) leads to the following formula, which is the basis for the so-called *optical flow* methods in computer vision [13]:

$$\frac{z}{k}\nabla\phi(\mathbf{r}) \cdot \nabla I_0(\mathbf{r}) + I(\mathbf{r}) - I_0(\mathbf{r}) = 0, \quad (3)$$

where \cdot denotes inner product. Note that the linearization is iteratively updated during the optimization, so that our model overall remains non-linear. A few numbers of iterative linearization suffices.

Compared to TIE that requires axial phase diversity, our model needs no z -direction displacements, because the distorted wavefront results in local transverse motion of the diffraction pattern. Our model Eq. (2) is nonlinear, involving first-order terms of the wavefront, whereas TIE is a linear model that employs second-order terms. Finally, in terms of optical path, Eq. (2) permits white light illumination, whereas TIE only allows for coherent or partially coherent illumination.

2.1. Sensor performance

The wavefront slope is determined by the following relationship:

$$\nabla\phi = a \frac{2\pi d_s}{\lambda z}, \quad (4)$$

where d_s is the sensor pixel size, and a is the amount of apparent motion measured in numbers of pixels. For a given tracking algorithm to solve Eq. (2), a limits itself to a certain range, e.g. $a \in [a_{\min}, a_{\max}]$. The sensitivity (or, the accuracy) of the sensor depends on a_{\min} , whereas the dynamic range depends on a_{\max} .

Denoting the local wavefront radius as R , one of the approximation made in deriving Eq. (1) requires (see Appendix):

$$\frac{z}{k} \nabla^2 \phi(\mathbf{r}) = \frac{z}{R} \approx 0, \quad (5)$$

which indicates $R \gg z$. For a given distance z , Eq. (5) can be used to estimate the maximum curvature $1/R$ that can be measured by our sensor.

2.2. Numerical solver

While the purpose of the Coded Wavefront Sensor is to estimate $\phi(\mathbf{r})$ from $I_0(\mathbf{r})$ and $I(\mathbf{r})$, standard optical flow methods seek to reconstruct the per-pixel apparent motion vectors between the reference image and measurement image, i.e. the gradient of the phase function. Instead of using standard optical flow algorithms like Horn-Schunck [13], we therefore devise our own reconstruction model that directly solves for the phase function itself and regularizes the apparent motion to be curl free.

Define the unknown wavefront as ϕ , the image gradient field as $(\mathbf{g}_x, \mathbf{g}_y) = \nabla I_0(\mathbf{r})$, and a “time” derivative $\mathbf{g}_t = I(\mathbf{r}) - I_0(\mathbf{r})$, at each linearisation, the reconstruction problem is formulated as:

$$\underset{\phi}{\text{minimize}} \quad \|\mathbf{G}\mathbf{M}\nabla\phi + \mathbf{g}_t\|_2^2 + \alpha \|\nabla\phi\|_2^2, \quad (6)$$

where $\mathbf{G} = [\text{diag}(\mathbf{g}_x) \ \text{diag}(\mathbf{g}_y)]$ is a concatenated diagonal matrix with the image derivatives on the diagonal, and \mathbf{M} is a binary diagonal matrix that selects only the M visible pixels from the N wavefront samples that affect the measurement, for direction x and y respectively.

Like existing optical flow methods, Eq. (6) contains both a “photoconsistency” term (first term) that describes the apparent motion of the pattern in the image plane, and a smoothness regularization term (second term), which is controlled with an additional parameter $\alpha > 0$. With a smoothness regularizer added directly to the phase function, Alternating Direction Method of Multipliers (ADMM) [15] is used to solve this optimization problem, where each updating step involves either element-wise operations or fast Fourier transforms (FFT), and hence is naturally parallelizable.

3. Implementation

3.1. Hardware

Our prototype Coded Wavefront Sensor (see Fig. 4) exploits a uniformly random binary mask pattern, and a GS3-U3-14S5M-C PointGrey monochromatic 2/3" CCD sensor with a pixel pitch of 6.45 μm. The mask is placed on top of the bare sensor, with a spacing of approximately 1.5 mm. The binary mask is fabricated using photolithography in a chrome layer deposited on a 4" Fused Silica wafer, with a pixel pitch of 12.9 μm.

3.2. Software

Our GPU version is implemented in CUDA, and the CuFFT library is used for all the involved FFT operations, which are the most time-consuming parts. To make the best utilization of CuFFT, the unknown width and height are set to be the powers of small primes, e.g. two or three.

4. Results

In this section the results are presented, including the synthetic ones in simulation, the experimental validation using a Spatial Light Modulator (SLM), and the realistic wavefront visualization of heat flow and defocusing.

All numerical experiments are run with a fixed regularization parameter. Both the CPU and GPU version of our algorithm are run on a workstation that has 125.8 GB RAM, exploits Intel Xeon E5-2697V3 @ 2.60 GHz \times 16 as CPU, and GeForce GTX TITAN X (Pascal) as GPU, with a Ubuntu 14.04 Linux as operating system.

4.1. Simulation

We have conducted two simulations to investigate the sensitivity and accuracy of the Coded Wavefront Sensor. In the simulations, the illumination is monochromatic (as of $\lambda = 550$ nm). The overall aperture size equals 6.6 mm \times 6.6 mm with sensor and mask pixel pitch be 6.45 μm and 12.9 μm respectively. The scalar field of interest is sampled with 1.29 μm . Gaussian noise is added and the image signal-to-noise-ratio (SNR) equals 40 dB. The wave propagation is simulated using the angular spectrum method [16] with filtering [17] to suppress high frequency artifacts.

The first numerical experiment evaluates the dynamic range of our sensor. A planar wave (i.e. the reference), and sixteen different scales of spherical waves are simulated at the mask plane, respectively, for five different distances z . The reference image, and sixteen measurement images are consequently recorded at the sensor plane. Figure 2 shows the wavefront reconstruction error (in terms of root mean square, RMS) our sensor can attain, providing the fixed curvature wavefronts that progressively violate Eq. (2). With the increase of wavefront range, the decrease of accuracy can be partially explained by the approximations made to derive Eq. (1).

Figure 3 shows the second numerical experiment, where we evaluate the performance of our sensor by sensing typical atmospheric turbulence. The same turbulence is evaluated at different scales. The synthetic atmospheric turbulence respects the Kolmogorov's theory, and is implemented using the sub-harmonic method [18]. The outer scale and inner scale of the base turbulence are set to be 4 m and 1 mm respectively. The mask-to-sensor distance $z = 1.5$ mm. The result indicates the possibility to apply our sensor for atmospheric turbulence measurement.

4.2. Quantitative experimental demonstration

We also evaluated the accuracy of Coded Wavefront Sensor by using it to measure known wavefronts. The experimental setup is shown in Fig. 4. The different target wavefronts are generated using a reflective LCoS-based SLM, the PLUTO SLM by HOLOEYE, whose pixel pitch is 8.0 μm , and the maximum phase retardation of 2π . We use phase wrapping to simulate larger phase retardations. This introduces higher diffraction orders in the synthesized waveforms, which are detected by the Coded Wavefront Sensor and show up as noisy structures in Fig. 5. This limits the experiments to small and medium distortions. We used different scales of cubic wavefronts, spherical wavefronts, customized wavefronts, and Zernike single-mode wavefronts as ground-truth targets.

A distant, white point light source is used as collimated illumination. A polarizer is placed in the optical path to produce linearly polarized illumination, which is required for the SLM to operate as a pure phase modulator. Behind the beamsplitter, a Kepler telescope structure is

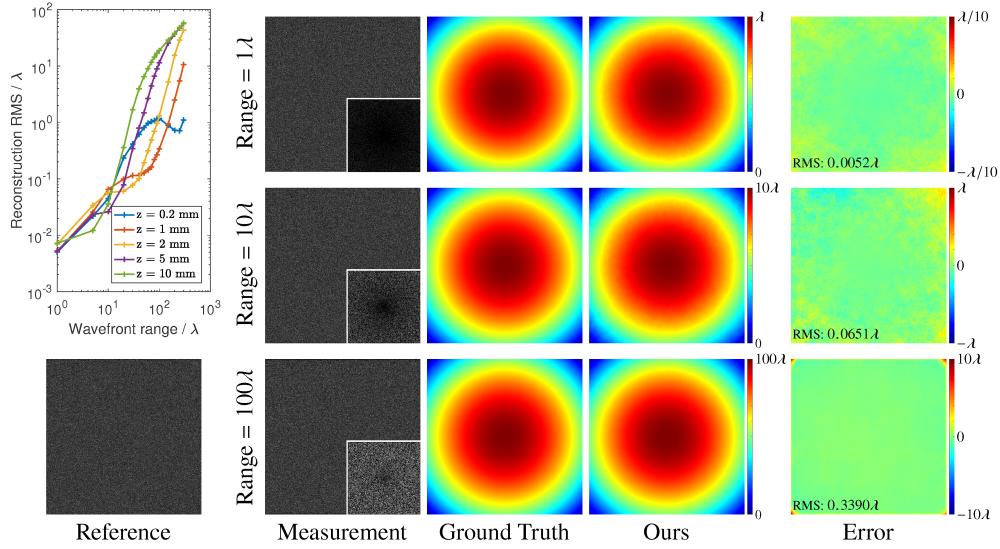


Fig. 2. Accuracy experiments. Top left shows the reconstruction error in RMS for different wavefront range (Peak-to-Valley) for different z . Specifically, results of three wavefront range spherical waves are shown when $z = 1 \text{ mm}$. To visualize the difference between the measurement and the reference, the logarithm of their subtraction are shown as inset.

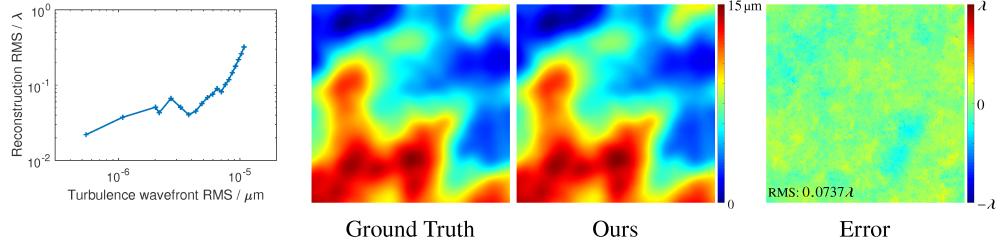


Fig. 3. Synthetic atmospheric turbulence. Most left shows the reconstruction RMS versus the turbulence RMS. Specifically, one scale of the turbulence is shown on the right.

imposed to ensure the SLM and our wavefront sensor are in conjugate planes, so the distorted wavefront measured by Coded Wavefront Sensor is the one produced by the SLM. In our setup, the Kepler telescope is composed of two lenses, with focus length $f_1 = 100 \text{ mm}$ and $f_2 = 75 \text{ mm}$ respectively. The sensor exposure time is set to be 5 ms at both the calibration step and the measurement step.

Selected experimental results for several different megapixel (1024×1024) wavefront shapes are visualized in Fig. 5 as synthetic interferograms (one interference ring equals $\lambda = 632.8 \text{ nm}$). The typical mean absolute phase error is below 50 nm, and is slightly larger than the ones in the simulation. We believe that the quantization and phase wrapping in the SLM that we use to generate the distortion are introducing extra errors. The ground truth wavefront is not really the actual wavefront distortion for that reason, and the isocontours in the recovered wavefront can be clearly seen. The RMS for all the experimental wavefronts are listed in Table 1, where for most wavefronts the reconstruction error is below 0.1 wavelengths.

We achieve a reconstruction speed of up to 50 frames per second, i.e. 50 million individual phase measurements per second. It is possible to increase the frame rate by sacrificing spatial resolution. Further improvements in speed should be possible through numerical strategies such

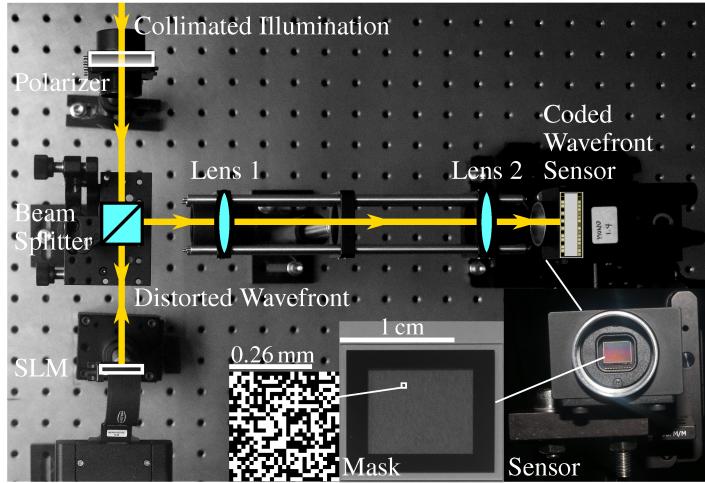


Fig. 4. Experimental setup for accuracy validation. Under collimated incoherent illumination from a white light source, the SLM generates a known distorted wavefront, which is then captured by our Coded Wavefront Sensor in the conjugate plane.

as “warm starting”, i.e. initializing the solution for one frame with the solution from the previous frame.

For reference, the GPU timing performance for different wavefront resolutions is in Table 2. All the timing experiments were run repeatedly for 1000 times, and the average is chosen as the final timing statistics. Our CUDA GPU version greatly improves the timing performance, compared to the CPU version that is implemented in MATLAB. For example, the average computing time for $N = 1024 \times 1024$ is 2.00 s in the CPU version, over which our GPU implementation achieves ~ 102 times speedup. For one megapixel wavefront resolution, we have achieved wavefront reconstruction speed as 51 Hz, which is much beyond real-time performance.

Table 1. Reconstruction error for all experimental wavefronts, where $\lambda = 632.8$ nm.

| Label | Range (in λ) | RMS (in λ) | Label | Range (in λ) | RMS (in λ) |
|---------------|-----------------------|---------------------|---------------|-----------------------|---------------------|
| Cubic #1 | 0.35 | 0.05 | Spherical #1 | 0.43 | 0.03 |
| Cubic #2 | 0.69 | 0.03 | Spherical #2 | 0.86 | 0.06 |
| Cubic #3 | 1.19 | 0.04 | Spherical #3 | 1.28 | 0.07 |
| Cubic #4 | 1.90 | 0.06 | Spherical #4 | 1.71 | 0.05 |
| Cubic #5 | 2.83 | 0.05 | Spherical #5 | 2.14 | 0.05 |
| Cubic #6 | 4.03 | 0.08 | Spherical #6 | 2.57 | 0.07 |
| Cubic #7 | 5.53 | 0.13 | Spherical #7 | 2.99 | 0.09 |
| Zernike m2 #1 | 1.52 | 0.09 | Spherical #8 | 3.42 | 0.12 |
| Zernike m2 #2 | 3.04 | 0.08 | Spherical #9 | 3.85 | 0.11 |
| Zernike m2 #3 | 6.07 | 0.08 | Zernike m5 #1 | 1.15 | 0.04 |
| Zernike m3 #1 | 1.52 | 0.05 | Zernike m5 #2 | 2.32 | 0.07 |
| Zernike m3 #2 | 3.04 | 0.07 | Zernike m5 #3 | 4.64 | 0.08 |
| Zernike m3 #3 | 6.07 | 0.11 | Zernike m6 #1 | 0.58 | 0.04 |
| Zernike m4 #1 | 1.16 | 0.05 | Zernike m6 #2 | 1.16 | 0.03 |
| Zernike m4 #2 | 2.32 | 0.07 | Zernike m6 #3 | 2.32 | 0.08 |
| Zernike m4 #3 | 4.64 | 0.15 | Customized #1 | 2.19 | 0.09 |
| Customized #2 | 4.38 | 0.06 | Customized #3 | 6.56 | 0.17 |

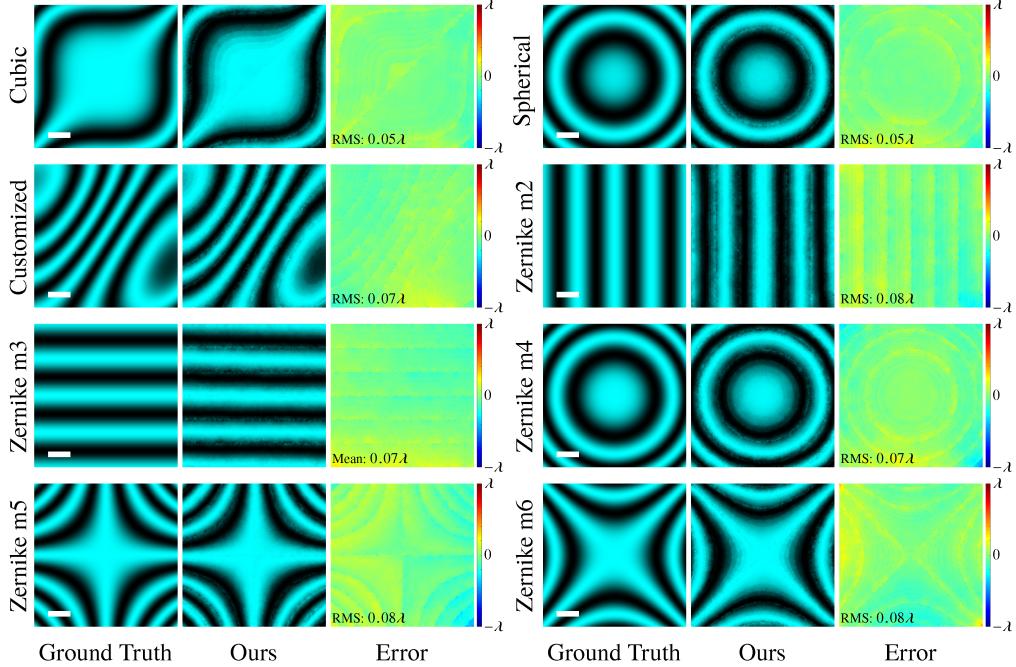


Fig. 5. Selected experimental results. The ground truth wavefronts, our reconstructed wavefronts, and the wavefront errors are shown. The wavefronts are shown in interference fringes where one fringe maps to wavefront difference of $\lambda = 632.8$ nm. Scale bar is 1 mm.

Table 2. Timing performance of our GPU implementation for solving Eq. (3).

| Unknown size | Performance |
|--------------------|-------------------------|
| 1024×1024 | 19.5 ms / 51.2 frames/s |
| 1024×768 | 15.2 ms / 65.6 frames/s |
| 640×480 | 11.0 ms / 91.2 frames/s |
| 512×512 | 5.8 ms / 173.3 frames/s |
| 256×256 | 3.0 ms / 339.0 frames/s |
| 128×128 | 2.2 ms / 460.8 frames/s |

4.3. Realistic wavefront imaging

We also visualized two realistic wavefronts, the ones created by heat flow and defocusing, using the Coded Wavefront Sensor. The heat flow is generated using a lighter, and defocus is achieved by manually moving a convex lens back and forth. To increase the field-of-view for better visualization, we employed a telescope system for wavefront magnification, with a ratio of two. All the measurement images were captured and the wavefronts were reconstructed on GPU in real-time. Please see [Visualization 1](#) for the experimental setup, the whole captured data and the reconstructed wavefronts. Here, two frames are chosen to be shown in Fig. 6, with their reconstructed wavefronts in interference fringes respectively.

5. Discussion

The Coded Wavefront Sensor is related to a number of other imaging systems and designs. Just like the Shack-Hartmann sensor can be interpreted as a combination of a lenslet-based light

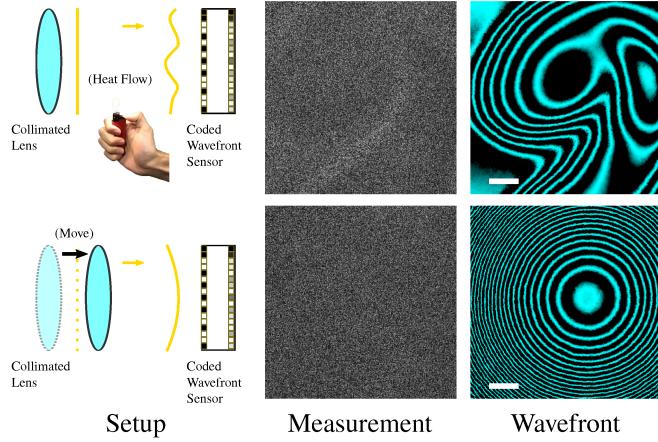


Fig. 6. Wavefront visualization of the heat flow and the defocusing. The setup diagrams are simplified versions of the real situations. See [Visualization 1](#) for full experimental details. Scale bar is 2 mm.

field camera [19] and 2D spot tracking software, our sensor can be seen as a combination of a mask-based light field camera [20, 21] with a more sophisticated dense 2D motion tracking method. The Coded Wavefront Sensor also bears similarity to Background Oriented Schlieren (BOS) imaging [22], but with the patterned “background” moved into the camera for a compact form factor.

The design of the Coded Wavefront Sensors allows it to be used as a drop-in replacement for any optical system currently using a Shack-Hartmann sensor, with an immediate gain in spatial resolution. In addition, we believe it can also be incorporated into optical systems configured for phase retrieval problems, including in microscopy. Since there is no need for phase diversity and coherent illumination, such an adaptation should in fact be easier than many existing phase retrieval setups. We will explore these applications in future works.

Several other extensions in both applications and usage are also conceivable in the future. The calibration or reference image does not always need to correspond to plane wave illumination, but could be a pre-distorted wavefront. For example when characterizing freeform lenses, a known ground-truth lens can be used to form the reference image, and the Coded Wavefront Sensor can then be used to characterize the difference between the reference lens and another lens.

For the accuracy of the Coded Wavefront Sensors it is necessary that the mask produces a locally distinctive diffraction pattern on the image sensor. To facilitate this process, the mask could be custom-designed (instead of random) to produce a specific diffraction pattern such as wavelet noise [23]. It should also be possible to use grayscale masks or even random phase gratings as an alternative to the binary masks employed in this work.

6. Conclusion

In conclusion, we introduce the Coded Wavefront Sensor, a novel sensor design that is physically implemented by a single binary masked sensor to encode the incoming wavefront, and is numerically implemented by an efficient optimization decoding algorithm, such that wavefront reconstruction with high spatio-temporal resolution is achieved within sub-wavelength accuracy. The theoretical principle behind Coded Wavefront Sensors offers a new approach to the wavefront sensing problem, namely the direct 2D tracking of diffraction patterns.

Appendix

Our mathematical model is based on modifications and simplification of the Green's function model by Teague [9] (Appendix A.3). Consider a general scalar field $u_0(\mathbf{r})$ (where $\mathbf{r} = (x, y)$ denotes a coordinate point on a 2D plane of interest) with wave number k , on the mask plane:

$$u_0(\mathbf{r}) = f_0(\mathbf{r})p_0(\mathbf{r}), \quad (7)$$

where $f_0(\mathbf{r}) = \exp[j\phi(\mathbf{r})]$ is the scalar field corresponding to the wavefront $\phi(\mathbf{r})$ that we want to measure, and $p_0(\mathbf{r})$ is the transfer function of the mask. We assume:

- The mask $p_0(\mathbf{r})$ is of high frequency (in our case, it is uniformly random binary), whose Fourier transform $P_0(\rho)$ (where ρ is the Fourier dual of \mathbf{r}) is broadband.
- The wavefront $\phi(\mathbf{r})$ is of low-frequency, and is second-order differentiable, and hence the distorted scalar field $f_0(\mathbf{r})$ is smooth enough such that its spectrum is bandlimited, and decays sufficiently close to zero in high frequency regions.

We now consider the wave propagation problem for small distance z . For a general mask $p_0(\mathbf{r})$, the Fresnel approximation may not be valid, and the more general Huygens-Fresnel based principle must be applied. Using the compact form of the Rayleigh-Sommerfeld diffraction formula, and expand it in the Fourier domain, the diffractive scalar field $u_z(\mathbf{r})$ is:

$$\begin{aligned} u_z(\mathbf{r}) &= \exp\left[jkz\left(1 + \frac{\nabla^2}{k^2}\right)^{1/2}\right]u_0(\mathbf{r}) \\ &= \int \exp(j2\pi\mathbf{r} \cdot \rho) \exp\left[jkz\left(1 - \lambda^2\|\rho\|_2^2\right)^{1/2}\right] \times \int P_0(\rho') F_0(\rho - \rho') d\rho' d\rho \\ &\approx \exp(-jkz) \int \exp(j2\pi\mathbf{r} \cdot \rho') \exp\left[jkz\left(1 - \lambda^2\|\rho'\|_2^2\right)^{1/2}\right] \times \\ &\quad \int \exp(j2\pi(\mathbf{r} - \lambda z \rho') \cdot \rho'') \exp\left[jkz\left(1 - \lambda^2\|\rho''\|_2^2\right)^{1/2}\right] F_0(\rho'') d\rho'' P_0(\rho') d\rho' \\ &= \exp(-jkz) \int \exp(j2\pi\mathbf{r} \cdot \rho') \exp\left[jkz\left(1 - \lambda^2\|\rho'\|_2^2\right)^{1/2}\right] \times \\ &\quad P_0(\rho') \exp\left[jkz\left(1 + \frac{\nabla^2}{k^2}\right)^{1/2}\right] f_0(\mathbf{r} - \lambda z \rho') d\rho', \end{aligned} \quad (8)$$

where the third equality of Eq. (8) results from the introduction of variable $\rho'' = \rho - \rho'$. And the approximation step comes from:

$$(1 - \lambda^2\|\rho\|_2^2)^{1/2} \approx (1 - \lambda^2\|\rho'\|_2^2)^{1/2} + (1 - \lambda^2\|\rho''\|_2^2)^{1/2} - \lambda^2\rho' \cdot \rho'' - 1, \quad (9)$$

which holds for smooth wavefront $F_0(\rho'')$ with $\lambda^2\|\rho''\|_2^2 \ll 1$. The last term in Eq. (8) is the diffraction of $f_0(\mathbf{r} - \lambda z \rho')$, for which the n th diffraction order $[(z/k)\nabla^2\phi(\mathbf{r})]^n = (z/R)^n \approx 0$ when local radius $R \gg z$. Also, a Taylor expansion of $\phi(\mathbf{r} - \lambda z \rho')$ up to second-order suggests:

$$\begin{aligned} \phi(\mathbf{r} - \lambda z \rho') &\approx \phi(\mathbf{r}) - \lambda z \rho' \cdot \nabla \phi(\mathbf{r}) + \frac{1}{2} \lambda z \rho' \cdot (\lambda z \nabla^2 \phi(\mathbf{r})) \rho' \\ &= \phi(\mathbf{r}) - \lambda z \rho' \cdot \nabla \phi(\mathbf{r}) + \pi \lambda z \frac{z}{R} \|\rho'\|_2^2. \end{aligned} \quad (10)$$

The linear approximation is reasonable when the second-order term approximates zero. It implies a lower bound d_m^* for the mask pixel pitch d_m , by the Nyquist Theorem that $\|\rho'\|_2^2 \leq 1/(4d_m^2)$:

$$d_m \gg d_m^* = \sqrt{\frac{\pi \lambda}{R}} \cdot \frac{z}{2}. \quad (11)$$

In our prototype, $d_m = 12.9 \mu\text{m}$ and $d_m^* \approx 1 \mu\text{m}$ for $R = 1 \text{ m}$. Given $R \gg z$ and the valid linear approximation of $\phi(\mathbf{r})$, we approximate the last term in Eq. (8) as:

$$\exp \left[jkz \left(1 + \frac{\nabla^2}{k^2} \right)^{1/2} \right] f_0(\mathbf{r} - \lambda z \boldsymbol{\rho}') \approx \exp(jkz) \exp[j\phi(\mathbf{r})] \exp[-j\lambda z \boldsymbol{\rho}' \cdot \nabla \phi(\mathbf{r})]. \quad (12)$$

Substituting Eq. (12) into Eq. (8), it yields:

$$u_z(\mathbf{r}) \approx \exp[j\phi(\mathbf{r})] p_z(\mathbf{r} - (z/k)\nabla\phi(\mathbf{r})), \quad (13)$$

where $p_z(\mathbf{r} - (z/k)\nabla\phi(\mathbf{r}))$ is the diffractive scalar field of $p_0(\mathbf{r})$ at the sensor plane.

Funding

King Abdullah University of Science and Technology (KAUST) baseline funding; KAUST Advanced Nanofabrication Imaging and Characterization Core Lab.

Acknowledgments

The authors would like to thank Yifan Peng for his help in the initial optical setup and valuable discussions.