

Accelerating a spatially varying aberration correction of holographic displays with low-rank approximation

SEUNG-WOO NAM,  DONGYEON KIM, AND BYOUNGHO LEE*

School of Electrical and Computer Engineering, Seoul National University, Gwanak-Gu Gwanakro 1, Seoul 08826, Republic of Korea

*Corresponding author: byoungcho@snu.ac.kr

Received 6 May 2022; revised 30 May 2022; accepted 30 May 2022; posted 31 May 2022; published 22 June 2022

Correction of spatially varying aberrations in holographic displays often requires intractable computational loads. In this Letter, we introduce a low-rank approximation method that decomposes sub-holograms into a small number of modes, thereby reformulating the computer-generated hologram calculation into a summation of a few convolutions. The low-rank approximation is carried out with two different algorithms: the Karhunen–Loeve transform as the optimum solution with respect to the mean-squared error criterion and a novel, to the best of our knowledge, optimization method to provide uniform image quality over the entire field of view. The proposed method is two orders of magnitude faster than the conventional point-wise integration method in our experimental setup, with comparable image quality.

2022 Optica Publishing Group under the terms of the [Optica Open Access Publishing Agreement](#)

<https://doi.org/10.1364/OL.462955>

Recent progress in near-eye displays aims to achieve a photorealistic image over a wide field of view for an immersive user experience [1,2]. Although a high numerical aperture eyepiece is necessary for a wide field of view, the optical aberration from the eyepiece gets severe as the numerical aperture increases. Spatially varying (SV) aberrations from non-ideal off-axis behavior of the high numerical aperture eyepiece hinder the realization of high-quality images over the wide field of view in near-eye displays. In this respect, holographic near-eye displays have been proposed, showing their capabilities in correcting optical aberrations [3–5]. Since the conventional point-wise integration method for SV aberration correction requires an intractable computational load, approximation methods such as spatially invariant (SI) correction, piece-wise correction, iterative optimization, and the deep learning-based method have been proposed [4–9]. However, a fast SV correction method with high reliability has not been introduced.

Here, we propose a low-rank approximation method that accelerates the calculation of a computer-generated hologram (CGH) for SV aberration correction. The proposed low-rank approximation method decomposes the sub-hologram into a small number of modes, approximating the point-wise integration as a summation of a few convolutions between the

target image and the modes, as shown in Fig. 1. In addition, we present two sub-hologram decomposition methods: the Karhunen–Loeve transform (KLT) and a uniform sub-hologram decomposition (USHD). The KLT is a decomposition method widely used in computer vision and computational imaging [10–12]. We use the KLT as a baseline method since it offers an optimum solution with respect to the mean-squared error (MSE) of the sub-hologram reconstruction. Furthermore, we propose the USHD as a novel decomposition method that optimizes the modes to have identical reconstruction errors at every target point. The USHD prevents the mode from being biased to the frequent sub-holograms and delivers the image with uniform quality over the entire field of view. We experimentally demonstrate the proposed method with a bench-top holographic display and analyze the results with the image contrast and the computation time.

SV aberration correction in holographic displays is performed by finding the phase profile of the sub-hologram that reconstructs a sharp point in the space. The point-wise integration method sums up all the sub-holograms corresponding to every point of the target [4]. In previous research, the phase profile of the sub-hologram is expressed with Zernike polynomials that are obtained with a manual calibration or optical simulation [5,6]. Therefore, a complex amplitude of CGH u_{sv} with a target amplitude a_{target} is expressed as

$$u_{sv}(x, y) = \iint h_{\tilde{x}, \tilde{y}}(x - \tilde{x}, y - \tilde{y}) a_{target}(\tilde{x}, \tilde{y}) d\tilde{x} d\tilde{y}, \quad (1)$$

where (x, y) is the CGH coordinate, (\tilde{x}, \tilde{y}) is the target image coordinate, and $h_{\tilde{x}, \tilde{y}}$ is the sub-hologram at (\tilde{x}, \tilde{y}) . Since CGH is calculated with quadruple loops over $\tilde{x}, \tilde{y}, x, y$, the time complexity of the point-wise integration method is $O(N_x N_y M_x M_y)$, where N_x, N_y and M_x, M_y are the width and height of the image and the sub-hologram, respectively. The computation time of the point-wise integration ranges from tens of seconds to tens of minutes for FHD images, even on high-performance graphics processing unit (GPU) [4,5].

SI correction uses a single sub-hologram for every target point and calculates the CGH with a convolution, neglecting the SV characteristic of the sub-holograms. Since the convolution can be replaced with a fast Fourier transform, the time complexity of the SI correction is $O(N_x N_y \log(N_x N_y))$, with an assumption of $N_x, N_y > M_x, M_y$. Although the SI correction significantly shortens

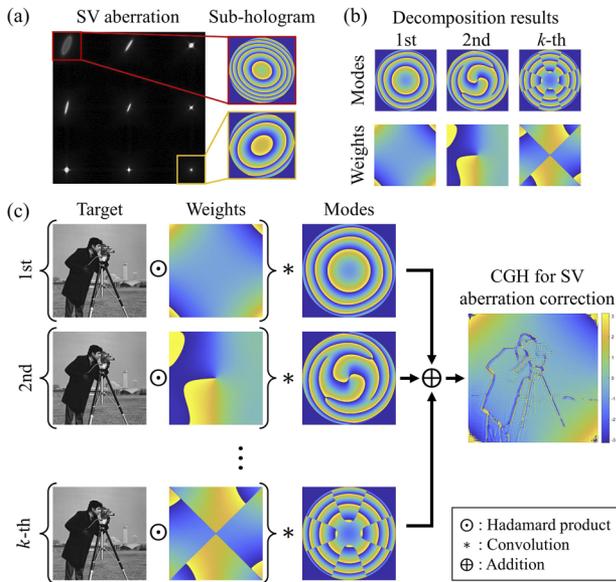


Fig. 1. Proposed low-rank approximation method. The phase values of the sub-holograms, modes, weights, and CGH are used for the visualization. (a) SV point spread functions and corresponding sub-holograms. (b) Sub-holograms decomposed into a small number of modes and weights. (c) CGH for SV correction obtained by a series of Hadamard products and convolutions.

the computation time, it only corrects the region near the point where the sub-hologram is selected.

The proposed low-rank approximation method decomposes the SV sub-holograms into a small number of modes with corresponding weights. Therefore, we can estimate the sub-hologram $h_{\tilde{x},\tilde{y}}$ as a weighted sum of the modes,

$$h_{\tilde{x},\tilde{y}}(x, y) \approx \sum_{m=1}^k r_m(\tilde{x}, \tilde{y}) v_m(x, y), \quad (2)$$

where k is a total number of the modes, m is an index of the modes, v_m is an m th mode, and r_m is a weight corresponding to v_m . We note that r_m is the SV weight in the target image domain, and v_m is the SI mode in the CGH domain. By substituting Eq. (2) into Eq. (1), u_{sv} is approximated as

$$u_{sv}(x, y) \approx \sum_{m=1}^k v_m(x, y) * (r_m(x, y) \cdot a_{\text{target}}(x, y)), \quad (3)$$

where $*$ denotes the convolution. The low-rank approximation reformulates the point-wise integration into k convolutions of the modes and the target image multiplied with the weights, as shown in Fig. 1(c). The time complexity of the low-rank approximation is $O(kN_x N_y \log(N_x N_y))$, which is only k times slower than the SI correction. Therefore, the low-rank approximation can reduce the computation time with a sufficiently small k .

A straightforward way for the low-rank decomposition is to minimize the total MSE of the reconstructed sub-holograms as

$$\min_{r,v} \sum_{\tilde{x},\tilde{y}} \left\| h_{\tilde{x},\tilde{y}} - \sum_{m=1}^k r_m(\tilde{x}, \tilde{y}) v_m \right\|_F^2, \quad (4)$$

where $\|\cdot\|_F$ denotes the Frobenius norm. A closed-form solution of the Eq. (4) is already known as the KLT [10,11,13]. Therefore,

we can obtain optimal sub-hologram modes by applying the KLT to the given sub-holograms. However, we observe that the KLT fails to correct the aberrations in the periphery of the image. As the aberration varies slowly near the optical axis, the KLT modes are biased to the central region to minimize the total MSE.

The USHD resolves this issue by forcing the reconstruction error to be identical regardless of the sub-hologram position. We optimize the modes to have identical reconstruction error \bar{e}_{KLT} using the objective function

$$\min_{r,v} \sum_{\tilde{x},\tilde{y}} \left\| h_{\tilde{x},\tilde{y}} - \sum_{m=1}^k r_m(\tilde{x}, \tilde{y}) v_m \right\|_F^2 - \bar{e}_{\text{KLT}}, \quad (5)$$

where \bar{e}_{KLT} is the average of the total MSE with the KLT modes v_{KLT} and weights r_{KLT} , which is expressed as

$$\bar{e}_{\text{KLT}} = \frac{1}{N} \sum_{\tilde{x},\tilde{y}} \left\| h_{\tilde{x},\tilde{y}} - \sum_{m=1}^k r_{\text{KLT},m}(\tilde{x}, \tilde{y}) v_{\text{KLT},m} \right\|_F^2. \quad (6)$$

Here, N is the total number of the sub-holograms, which is identical to $N_x N_y$. Equation (5) forces every reconstruction error to be \bar{e}_{KLT} identically, while the total MSE is equal to that of the KLT. In addition, we constrain the modes v to be orthonormal to each other to express the weight r_m as the inner product of the modes v_m and the sub-hologram $h_{\tilde{x},\tilde{y}}$. We solve the given constrained optimization problem with projected gradient descent (PGD), which performs alternations of gradient descent and projection to a feasible domain [14]. Here, PGD repeats the update of the modes via stochastic gradient descent followed by orthonormalization of the modes using QR decomposition.

We demonstrate the proposed method with a bench-top holographic display, as shown in Fig. 2(a). The display system consists of a laser diode (Thorlabs, CPS635F), a spatial light modulator (Thorlabs, EXULUS-HD1), a linear polarizer LP, collimating lens L1, and a 4- f system L2–L3. High-order and DC noises are eliminated with the spatial filter in the 4- f system. In addition, we adopt a tilted and shifted lens L4 with a 30-mm focal length to introduce SV aberrations. The captured image of point spread functions without aberration correction is shown in Fig. 2(b). The phase profile of the sub-hologram for SV correction is measured with a manual calibration. The sub-hologram with a diameter of 450 pixels is expressed with six Zernike polynomials: 4th, 5th, 6th, 7th, 8th, and 11th in the Noll index [15]. Both KLT and USHD algorithms are implemented with PyTorch, and the CGH generation part is implemented with CUDA C on a Nvidia RTX 3080Ti GPU. Furthermore, the amplitude discard method is used to encode phase-only CGH from the complex amplitude. In the USHD, we optimize the

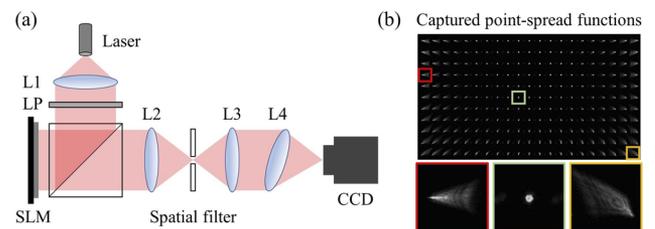


Fig. 2. (a) Diagram of the bench-top holographic display. SLM, spatial light modulator; L1, collimating lens; LP, linear polarizer; L2–L3, 4- f filtering system; L4, tilted and shifted lens to introduce SV aberration. (b) A captured image of point spread functions without aberration correction.

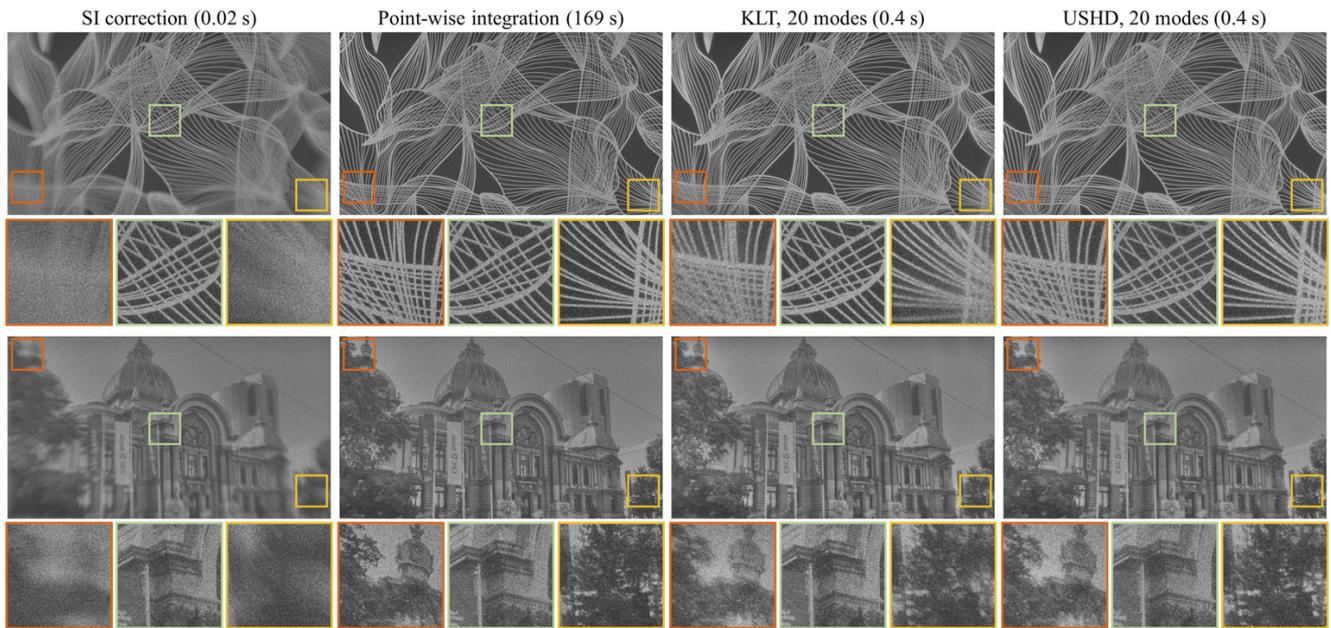


Fig. 3. Experimentally captured results with various aberration correction methods. The numbers in the parentheses indicate computation time for a single CGH with corresponding methods. The results of SI correction demonstrate that our system has severe SV aberration, which is corrected by the point-wise integration method. We can see the presence of the aberration in the periphery of the KLT results while the USHD with the same number of modes corrects both central and peripheral aberration successfully.

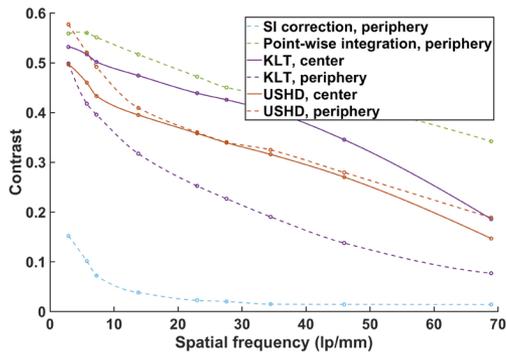


Fig. 4. MTF curves measured at the center and the periphery of the image with the various aberration correction methods. The number of modes used in KLT and USHD is both 20.

modes for 50,000 iterations with 150 batches. The stochastic gradient descent is performed with an ADAM optimizer in PyTorch with a learning rate of 0.0001.

Figure 3 shows experimental results of the various aberration correction methods in our holographic display setup. We digitally combined the captured images reconstructed from 25 CGHs with different random phases to reduce the speckle noise and emphasize the impact of the aberration in the image. Although the runtime of the SI correction is the shortest, 0.02 seconds, it suffers from image quality degradation since the aberration still remains even after the correction. In contrast, the point-wise integration method has the longest runtime, 169 seconds, with the best image quality. The proposed low-rank approximation method shows the results that are in between the SI correction and the point-wise integration method. The runtime of the low-rank approximation methods is 0.4 seconds, which is approximately 420 times faster than the point-wise integration

method, while the results are comparable. However, we can see from the result that KLT still shows image quality degradation in the peripheral region. The USHD with the identical number of modes successfully corrects the aberration in the peripheral region with a slight sacrifice of the image quality in the center.

We compare the modulation transfer function (MTF) of the KLT and the USHD in Fig. 4 to quantitatively show that the USHD prevents the biasing of the modes. The MTF is obtained by the intensity of 200×200 holographic images of binary gratings, cropped at the center and the bottom right. From the maximum intensity I_{\max} and minimum intensity I_{\min} of the cropped images, the contrast is calculated by $(I_{\max} - I_{\min}) / (I_{\max} + I_{\min})$. Figure 4 shows the MTF curves at the center and the periphery with the various aberration correction methods. Both KLT and USHD show higher image contrast compared to the SI correction in the periphery. However, the MTF of KLT at the center and the periphery largely differ because of the mode biasing, as we discussed before. In contrast, USHD shows similar MTF curves for both the center and periphery, placed between two MTF curves of the KLT. We note here that the overall low image contrast is due to the lack of an optimization process for phase-only encoding during the CGH generation. It can be further improved by combining the proposed method with conventional CGH optimization methods [16,17].

In the low-rank approximation method, the image quality increases and the calculation gets slower as the number of modes increases, and vice versa. The experimental result of the low-rank approximation method with a different number of modes is shown in Fig. 5. Both KLT and USHD correct the aberration better when more modes are used, while the runtime increases linearly with an increasing number of modes. MTF curves in Fig. 6(a), drawn with KLT and USHD with a different number of modes in the periphery, also show that the increased number of modes leads to a higher image contrast in both methods.

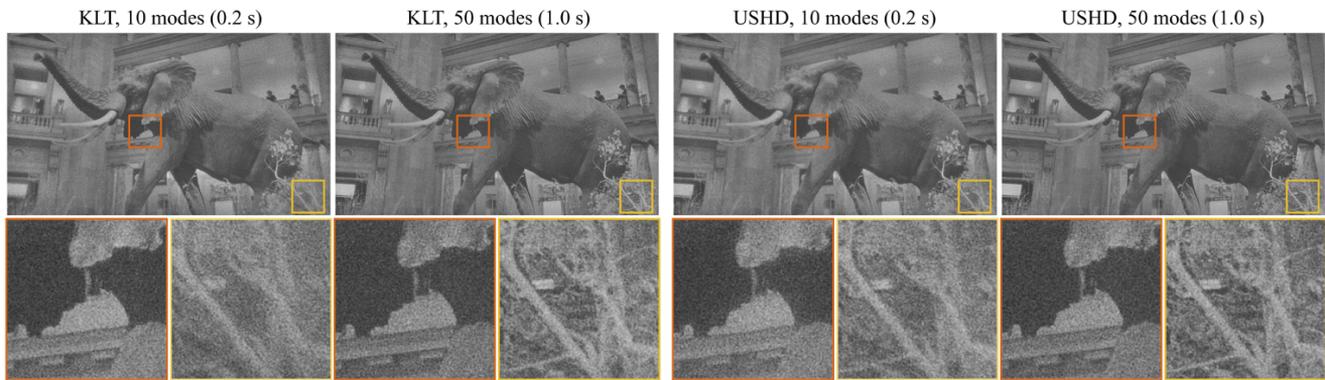


Fig. 5. Experimentally captured holographic images with a different number of modes in KLT and USHD. CGH calculation with a large number of modes corrects the aberration better in both methods. The KLT results only get better in the periphery since the central region is already corrected with only 10 modes. USHD results show image quality improvement at both the center and periphery. The numbers in the parentheses indicate the computation time for a single CGH.

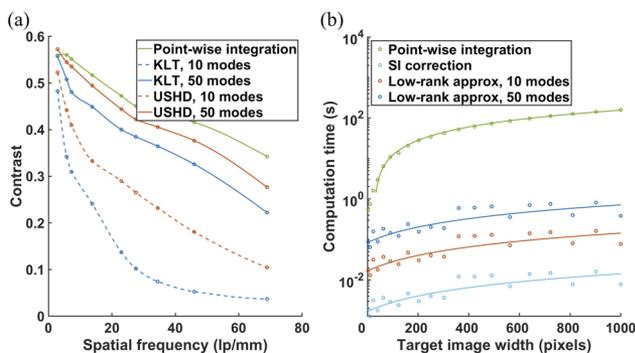


Fig. 6. (a) MTF curves of KLT and USHD with 10 and 50 modes, all in the periphery. The contrast gets higher as the number of modes increases. (b) Computation time versus the size of the target image, measured on the GPU. The solid lines represent fitted curves with corresponding time complexity. The low-rank approximation significantly reduces the computation time compared to the point-wise integration method.

Figure 6(b) shows the computation time of CGH versus the size of the target image, measured on the GPU. The diameter of the sub-hologram is 450 pixels, the same as in the experiment, and the width and the height of the target image are identical. Solid lines in the graph indicate the fitted curve with the corresponding time complexity; $O(N_x N_y M_x M_y)$ for the point-wise integration and $O(k N_x N_y \log(N_x N_y))$ for the others. We can see that the computation time of the low-rank approximation method is linearly proportional to the number of modes and far shorter than the point-wise integration method.

In this work, we present the low-rank approximation method for SV aberration correction in holographic displays, which accelerates the conventional point-wise integration method by 420 times without noticeable artifacts in the image. Also, the proposed USHD method provides an additional option of improving the image quality in the periphery with a slight sacrifice in the center, compared to the KLT. Although we demonstrate our method with a monochromatic holographic display, our method can be adopted to a full-color display setup easily by treating each color channel separately. In addition, our results suffer from low contrast due to the amplitude discard method used in the phase-only encoding. Future works could

improve the contrast of the image by plugging our method into iterative CGH optimization methods and replacing the wavefront propagation part during the CGH generation.

Funding. Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korean government (MSIT) (2021-0-00091).

Acknowledgment. This work was supported by Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korean government (MSIT) (No. 2021-0-00091, Development of real-time high-speed renderer technology for ultra-realistic hologram generation).

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

REFERENCES

1. K. Yin, Z. He, K. Li, and S.-T. Wu, *Opt. Express* **29**, 11512 (2021).
2. K. Bang, Y. Jo, M. Chae, and B. Lee, *IEEE Trans. Visual. Comput. Graphics* **27**, 2545 (2021).
3. D. Lee, K. Bang, S.-W. Nam, B. Lee, D. Kim, and B. Lee, *Sci. Rep.* **12**, 6649 (2022).
4. A. Maimone, A. Georgiou, and J. S. Kollin, *ACM Trans. Graph.* **36**, 1 (2017).
5. D. Kim, S.-W. Nam, K. Bang, B. Lee, S. Lee, Y. Jeong, J.-M. Seo, and B. Lee, *Biomed. Opt. Express* **12**, 5179 (2021).
6. A. Kaczorowski, G. S. D. Gordon, and T. D. Wilkinson, *Opt. Express* **24**, 15742 (2016).
7. S.-W. Nam, S. Moon, B. Lee, D. Kim, S. Lee, C.-K. Lee, and B. Lee, *Opt. Express* **28**, 30836 (2020).
8. T. Haist, A. Peter, and W. Osten, *Opt. Express* **23**, 5590 (2015).
9. D. Yoo, S.-W. Nam, Y. Jo, S. Moon, C.-K. Lee, and B. Lee, *J. Opt. Soc. Am. A* **39**, A86 (2022).
10. L. Denis, E. Thiébaud, F. Soulez, J.-M. Becker, and R. Mourya, *Int. J. Comput. Vis.* **115**, 253 (2015).
11. F. Sroubek, J. Kamenicky, and Y. M. Lu, *IEEE Signal Process. Lett.* **23**, 346 (2016).
12. K. Yanny, K. Monakhova, R. W. Shuai, and L. Waller, *Optica* **9**, 96 (2022).
13. A. Levey and M. Lindenbaum, *IEEE Trans. on Image Process* **9**, 1371 (2000).
14. D. P. Bertsekas, *J. Oper. Res. Soc.* **48**, 334 (1997).
15. R. J. Noll, *J. Opt. Soc. Am.* **66**, 207 (1976).
16. P. Chakravarthula, E. Tseng, T. Srivastava, H. Fuchs, and F. Heide, *ACM Trans. Graph.* **39**, 1 (2020).
17. Y. Peng, S. Choi, N. Padmanaban, and G. Wetzstein, *ACM Trans. Graph.* **39**, 1 (2020).