https://doi.org/10.2352/lim.2022.1.1.09

Effect of bit-depth in stochastic gradient descent performance for phase-only computer-generated holography displays

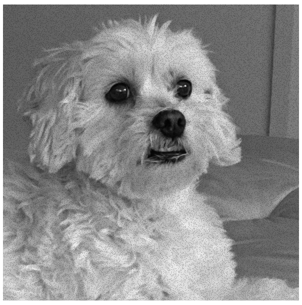
Andrew Kadis, Benjamin Wetherfield, Jinze Sha, Fan Yang, Youchao Wang, Timothy D. Wilkinson; University of Cambridge, Cambridge, UK

Abstract

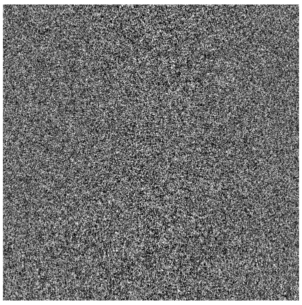
SGD (Stochastic gradient descent) is an emerging technique for achieving high-fidelity projected images in CGH (computer-generated holography) display systems. For real-world applications, the devices to display the corresponding holographic fringes have limited bit-depth depending on the specific display technology employed. SGD performance is adversely affected by this limitation and in this piece of work we quantitatively compare the impact on algorithmic performance based on different bit-depths by developing our own algorithm, Q-SGD (Quantised-SGD). The choice of modulation device is a key decision in the design of a given holographic display systems and the research goal here is to better inform the selection and application of individual display technologies.

Introduction

Computer-generated holography (CGH) is a family of techniques whereby the desired optical wavefront is generated by computing the corresponding holographic fringe patterns digitally using a computer [1]. Compared to traditional film-holography, where a given object's interference pattern must be optically recorded, CGH allows an object to be projected using purely synthetic data [2].



(a) CGH Replay field



(b) Corresponding phase hologram

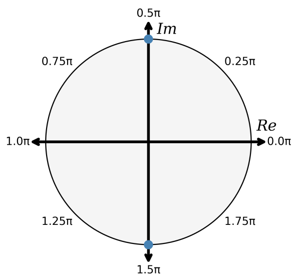
Figure 1: Monochrome CGH image generated using SGD

CGH forms the basis of holographic-displays, an emerging display technology where holographic approaches are used to recreate the corresponding optical wavefronts which would be observed in a real-world scene [3]. A perfect holographic display is capable of faithfully reproducing the corresponding depth cues from a physical object, an approach closer-suited to the human psycho-visual system than traditional displays [4, 5, 6]. Although the system-level challenges are considerable [7], significant advances have been made in this field the past few years, with a number of real-world systems delivering impressive results in both near-eye [8, 9, 10, 11, 12] and far-field display applications [13, 14, 15, 16, 17].

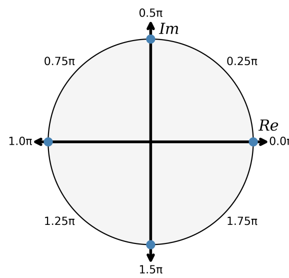
The key research challenge in CGH is how to compute the set of holographic fringes which correspond to a desired wavefront. There exists a large number of established algorithmic approaches to solve this problem, such as direct-search [18], phase-retrieval algorithms [19, 20], simulated-annealing [21], noise reduction time-multiplexing [22], double-phase methods [23], hardware feedback [24, 25], frequency-domain approaches[26] and formal optimisation [27]. The exact choice depends on the specific optical requirements, display device type selected and the specific computational hardware available [28, 29, 30].

An emerging approach is SGD (stochastic gradient descent), which has been shown to produce high-fidelity images for holographic display applications [31, 32, 33]. This is an iterative approach whereby the errors in the display output are optimised by taking successive steps along the slope of the output gradient for each given iteration. Although not a recently development, with the original work dating back to the 1950s [34], the advent of high-performance tools developed for Machine-Learning applications, such as *PyTorch* [35] with its in-built GPU acceleration and powerful automated differentiation capabilities [36], have enabled contemporary researchers to obtain high-quality results with SGD.

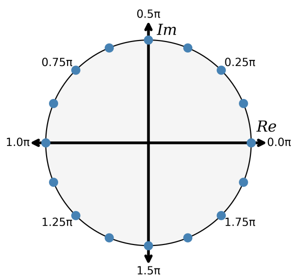
SGD for CGH display applications, an example of which is shown in Figure 1, is a *state-of-the-art* technique. However, one of the underlying assumptions of the mathematics behind SGD is that the input, in this case our display device, is continuous. The devices to display the corresponding holographic fringes have finite bit-depth dependent on the specific display technology employed. This quantisation error leads to a reduction in SGD performance as it is not possible for our device to directly take the steepest gradient-descent path.



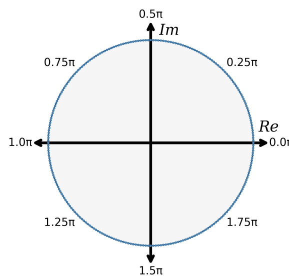
(a) Binary-phase encoding



(b) Two-bit encoding



(c) Four-bit encoding



(d) Eight-bit encoding

Figure 2: Hologram bit-depth is constrained by display type

Depending on the specific display technology employed, a typical display system can have multiple different display depths; one-bit [37], four-bit [38] and eight-bit[39] levels of modulation are all available with contemporary devices. These different modulation-level capabilites are shown overlaid on the complex-plane in Figure 2. The SGD algorithm must quantise the desired position for the subsequent iteration of gradient-descent. This has a detrimental effect on algorithm performance.

In this piece of work, we set-out to quantify the impact of this error, and hence understand the effect of bit-depth on SGD performance. This is an important question, as the bit-depth available is a direct function of the modulation technology selected for a given holographic display. It is hoped that the research presented here shall better inform the selection of appropriate technologies for a given application.

Research Objective

The objective of this research is to quantify the impact of different bit-depths on the SGD algorithm for CGH. In order to achieve this, we adopt the following approach:

1. We approach the problem in a manner which is agnostic to any specific display implementation, using simulation to explore the impact of the different levels of quantisation.
2. We develop our own algorithm, Quantised Stochastic Gradient Descent (Q-SGD), which extends the SGD algorithm such that it can to executed on an arbitrary bit-depth display device. Q-SGD iterates on previously developed SGD [31] (continuous phase) and B-SGD [40] (binary-amplitude) algorithms for high-fidelity holographic display images.
3. We compare the resultant output images from Q-SGD at a number of different bit-levels; binary-phase encoding, two-bit encoding, four-bit encoding and eight-bit encoding. Additionally we compare with a continuous-phase SGD to provide a control to benchmark against.

Implementation

To investigate the effect of bit-depth, we implemented the Q-SGD optimisation loop shown in Figure 3. A target replay field of amplitudes was used as the set-point, alongside an initial phase hologram of random phases, and iteratively optimised to produce the desired output.

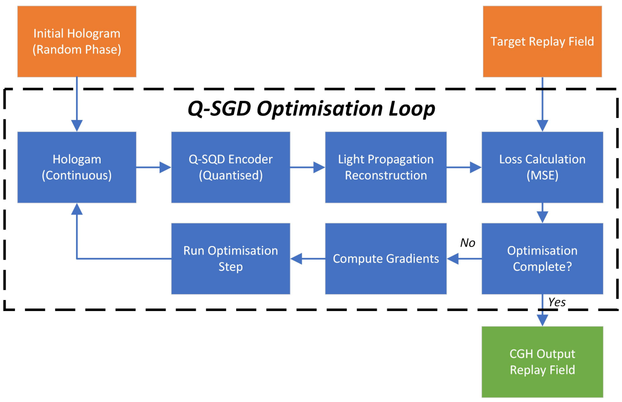


Figure 3: Q-SGD optimisation

For the light propagation simulation, we used the Band-Limited Angular Spectrum Method [41]. The optimisation loop was implemented using *PyTorch* configured to use the Adam optimiser [42]. Mean squared Error (MSE) between the target replay field and the amplitude of the propagation output was used as the optimisation metric.

Q-SGD Encoder

The key innovation in our implementation is the development of the Q-SGD algorithm. The novelty here is the usage of an arbitrary bit-depth encoder to allow the phase hologram to be quantised with N bits of depth to provide Q quantisation levels.

SGD requires the output to be a continuous mathematical function in order to be able to calculate the derivative, and hence the resultant gradient, for each given point. Hence it is not possible to use standard techniques for transforming continuous-phase to quantised-phase holograms. Peng et al. used continuous-phase holograms for the SGD algorithm [31] whilst Lee et al. [40] developed B-SGD in order to provide binary-amplitude quantisation by using a *HardTanh* function to provide binary quantisation whilst maintaining mathematical continuity in the output.

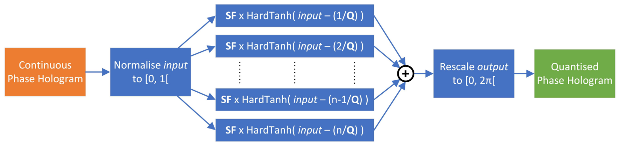


Figure 4: Q-SGD Encoder for Q levels of a given phase hologram

Our work here is a more general extension to B-SGD. Q-SGD, shown in Figure 4 and detailed in pseudo-code as Algorithm 1, uses multiple *HardTanh* functions shifted across the input's window to encode an arbitrary bit-depth. This allows the SGD family of approaches to be applied directly to a range of real-world devices with corresponding bit depths.

Algorithm 1: Q-SGD Encoding Algorithm

Input: Continuous Hologram to Encode H_i
Quantisation Bit-Depth Q_N
Scaling-Factor $scaling\_factor$
Output: Encoded Quantised Hologram H_e

```
// Normalise input to [0, 1.0[
```
H_n = Normalise(H_i)
```
// Work out the width of each bit
```
$Width_b$ = CalcBitWidth(Q_N)
```
// Calculate HardTanh steepness
```
SF = 1.0 / ($scaling\_factor$*$Width_b$)
```
// Iterate though Quantisation levels
```
for $n = 1$ **to** Q_N **do**
```
    // Sequentially shift along x-axis
```
    H_{Q0} = SF * HardTanh(H_n - (1.0-(n*$Width_b$)))
```
    // Summate our cumulative shifts
```
    H_{Q0} = H_{out} + H_{Q0}
```
// Rescale to original range and return
```
H_e = Rescale(H_{out})

In developing Q-SQD, it was observed that a key hyperparameter parameter was the slope of the *HardTanh* function. If it is too narrow, there is insufficient differentiation information passed through to the optimiser to effectively solve for the output. If it is too wide, the solver is effectively acting like a solver on a continuous set of inputs. It was found that setting the solver to a static value across all different bit-depths was overly harsh to the lower bit-depths as they were disproportionately penalised. Hence, the slope of the *HardTanh* function is defined as a percentage of bit-depth, *scaling_factor*. Empirically, it was found that setting *scaling_factor* to be approximately 20% of the range for each given bit-depth yielded good algorithmic performance whilst maintaining quantisation functionality.

Results

We executed the Q-SGD algorithm by running the *Mandrill* test image at binary, two-bit, four-bit, eight-bit and continuous-phase levels of modulation. The algorithm was executed for 200 iterations for all bit-depths; the propagation distance was set to 20 cm and the propagation was calculated for monochrome light at 638 nm with a SLM pixel-pitch of 6.4 μm. The implementation was executed in *Python 3.9.13* and *PyTorch 1.11.0* and an array of random phases was used as the initial hologram. The outputs are shown in Figure 6 with the errors captured in Table 1.

Table 1: Q-SGD performance at different levels of quantisation after 200 iterations; output fidelity is captured through both error (using MSE) and perceived-quality (using SSIM)

Modulation Bit-Depth	MSE	SSIM
Binary-Phase Q-SGD	0.0323	0.1949
Two-Bit Q-SGD	0.0457	0.1203
Four-Bit Q-SGD	0.0315	0.2113
Eight-Bit Q-SGD	0.0244	0.2650
Continuous-Phase SGD	0.0072	0.5358

Subjectively, the results are consistent with what is expected. A general trend of higher bit-depths leading to a lower absolute error, measured through MSE, and a higher degree of perceived quality, seen through higher structural similarity (SSIM) is evident.

The coarse structure of the target replay field test image is well-reconstructed. However there is a notable reduction in the fidelity of the high-frequency information, particularly fine-details such as the fur and texture; this is typical for lower bit-depth holograms. There is also a reduction in peak intensity at lower bit-depths. As expected, the continuous-phase image performs the best, with the a reduction in the produced image quality as the bit-depth decreases. The observed behaviour is largely along the lines of what is expected, with the exception of two-bit Q-SQD performance which is discussed below.

The convergence behaviour of the algorithm is plotted in Figure 6. For all bit-depths, a rapid reduction in MSE is observed, with the improvements plateauing to a constant error level after 50-100 iterations. One notable feature is the presence of *'knees'* in the solver (most pronounced in 4-bit Q-Bit Q-SGD) where the Q-SGD algorithm is able to descend down a rapid slope and large improvements in MSE are seen in a few iterations. Again, the expected general trend of higher bit-depths corresponding to higher performance levels is observed.

The exception to this is two-bit encoding which has considerably worse Q-SGD performance than the single-bit algorithm. There is an open question here of whether this is some artifact of our simulation or a reproducible limitation in the convergence behaviour of the gradient solver. Gradient-Descent solvers are typically optimised for linear systems and it is conceivable that the fourier transform in the light-propagation model may lead to reduced performance compared to single-bit. Alternatively, these solvers are susceptible to being caught in local minima and hence it is possible that the two-bit Q-SGD has become trapped at a lower-quality local minima than the other quantisation levels; alternative initial conditions and optimsiation techniques may alleviate this issue. Future work shall seek to replicate these results experimentally and seek to better understand the underlying behaviour.

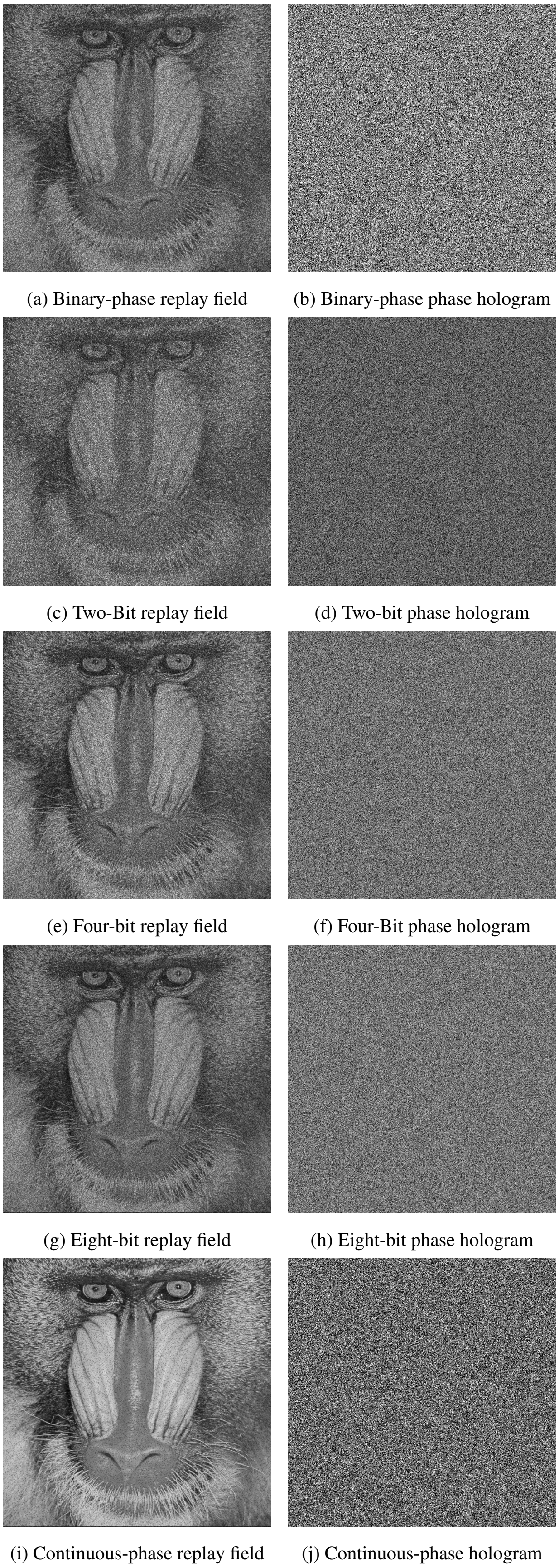
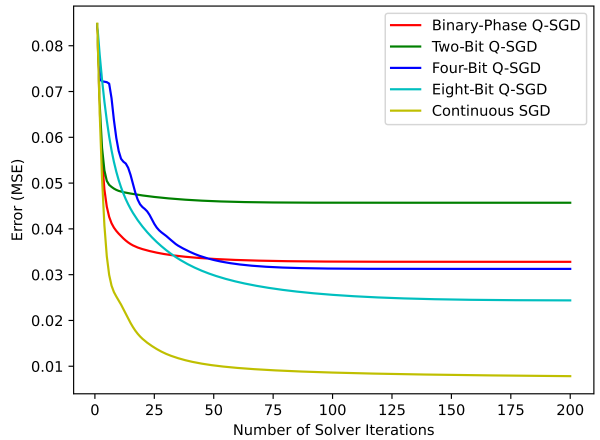
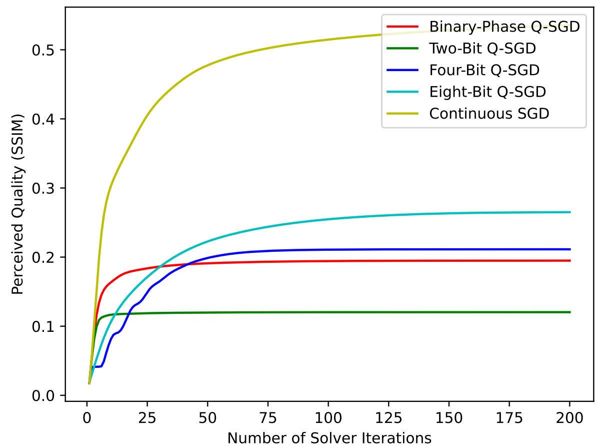


Figure 5: Simulated display outputs of Q-SGD and associated phase hologram at different bit-depths for *Mandrill* test image



(a) CGH error (as MSE) plotted against iteration number



(b) CGH perceived quality (as SSIM) plotted against iteration number

Figure 6: Q-SGD optimisation performance

Conclusion

In this piece of work, we have attempted to quantify the extent to which SGD can scale to different bit-depths. We have developed our own approach, Q-SGD which allows us to perform SGD with an arbitrary quantisation levels. We have successfully demonstrated that it can work for a number of arbitrary bit-depths and have quantified the algorithm performance.

The work presented here has been performed in simulation. As a follow-up, we shall attempt to reproduce the results experimentally and seek to better understand the impact of the slope of the *HardTanh* used in Q-SGD.

Acknowledgments

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC).

References

[1] Sahin, E., Stoykova, E., Mäkinen, J., and Gotchev, A. (2020) Computer-Generated holograms for 3D imaging: A survey. *ACM Comput. Surv.*, **53**, 1–35.

[2] Park, J.-H. (2017) Recent progress in computer-generated holography for three-dimensional scenes. *J. Soc. Inf. Disp.*, **18**, 1–12.

[3] Slinger, C., Cameron, C., and Stanley, M. (2005) Computer-generated holography as a generic display technology. *Computer*, **38**, 46–53.

[4] Hoffman, D. M., Girshick, A. R., Akeley, K., and Banks, M. S. (2008) Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. *J. Vis.*, **8**, 33–33.

[5] Chang, C., Bang, K., Wetzstein, G., Lee, B., and Gao, L. (2020) Toward the next-generation VR/AR optics: a review of holographic near-eye displays from a human-centric perspective. *Optica*, **7**, 1563–1578.

[6] Padmanaban, N., Konrad, R., Stramer, T., Cooper, E. A., and Wetzstein, G. (2017) Optimizing virtual reality for all users through gaze-contingent and adaptive focus displays. *Proc. Natl. Acad. Sci. U. S. A.*, **114**, 2183–2188.

[7] Blinder, D., Ahar, A., Bettens, S., Birnbaum, T., Symeonidou, A., Ottevaere, H., Schretter, C., and Schelkens, P. (2019) Signal processing challenges for digital holographic video display systems. *Signal Processing: Image Communication*, **70**, 114–130.

[8] Lim, Y., et al. (2016) 360-degree tabletop electronic holographic display. *Opt. Express*, **24**, 24999–25009.

[9] Li, G., Lee, D., Jeong, Y., Cho, J., and Lee, B. (2016) Holographic display for see-through augmented reality using mirror-lens holographic optical element. *Opt. Lett.*, **41**, 2486–2489.

[10] Peng, H., Cheng, D., Han, J., Xu, C., Song, W., Ha, L., Yang, J., Hu, Q., and Wang, Y. (2014) Design and fabrication of a holographic head-up display with asymmetric field of view. *Appl. Opt.*, **53**, H177–85.

[11] Shi, L., Huang, F.-C., Lopes, W., Matusik, W., and Luebke, D. (2017) Near-eye light field holographic rendering with spherical waves for wide field of view interactive 3D computer graphics. *ACM Trans. Graph.*, **36**, 1–17.

[12] Maimone, A., Georgiou, A., and Kollin, J. S. (2017) Holographic near-eye displays for virtual and augmented reality. *ACM Trans. Graph.*, **36**, 1–16.

[13] Oikawa, M., Shimobaba, T., Yoda, T., Nakayama, H., Shiraki, A., Masuda, N., and Ito, T. (2011) Time-division color electroholography using one-chip RGB LED and synchronizing controller. *Opt. Express*, **19**, 12008–12013.

[14] Kozacki, T., Finke, G., Garbat, P., Zaperty, W., and Kujawińska, M. (2012) Wide angle holographic display system with spatiotemporal multiplexing. *Opt. Express*, **20**, 27473–27481.

[15] Inoue, T. and Takaki, Y. (2015) Table screen 360-degree holographic display using circular viewing-zone scanning. *Opt. Express*, **23**, 6533–6542.

[16] Sasaki, H., Yamamoto, K., Wakunami, K., Ichihashi, Y., Oi, R., and Senoh, T. (2014) Large size three-dimensional video by electronic holography using multiple spatial light modulators. *Sci. Rep.*, **4**, 6177.

[17] Yaraş, F., Kang, H., and Onural, L. (2011) Circular holographic video display system. *Opt. Express*, **19**, 9147–9156.

[18] Jennison, B. K., Allebach, J. P., and Sweeney, D. W. (1989) Iterative approaches to Computer-Generated holography. *Organ. Ethic.*, **28**, 629–637.

[19] Gerchberg, R. W. (1972) A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik*, **35**, 237–246.

[20] Fienup, J. R. (1982) Phase retrieval algorithms: a comparison. *Appl. Opt.*, **21**, 2758–2769.

[21] Yang, H.-J., Cho, J.-S., and Won, Y.-H. (2009) Reduction of reconstruction errors in kinoform CGHs by modified simulated annealing algorithm. *J. Opt. Soc. Korea, JOSK*, **13**, 92–97.

[22] Cable, A. J., Buckley, E., Mash, P., and others (2004) 53.1: Real-time binary hologram generation for high-quality video projection applications. *Symposium Digest of . . . .*

[23] Sui, X., He, Z., Jin, G., Chu, D., and Cao, L. (2021) Band-limited double-phase method for enhancing image sharpness in complex modulated computer-generated holograms. *Opt. Express*, **29**, 2597–2612.

[24] Chakravarthula, P., Tseng, E., Srivastava, T., Fuchs, H., and Heide,

F. (2020) Learned hardware-in-the-loop phase retrieval for holographic near-eye displays. *ACM Trans. Graph.*, **39**, 1–18.

[25] Kadis, A., Mouthaan, R., Dong, D., Wang, Y., Wetherfield, B., El Guendy, M., and Wilkinson, T. D. (2021) Binary-Phase Computer-Generated holography using hardware-in-the-loop feedback. *OSA Imaging and Applied Optics Congress 2021 (3D, COSI, DH, ISA, pcAOP)*, Jul., p. DW5E.1, Optical Society of America.

[26] Wetherfield, B., Kadis, A., and Wilkinson, T. D. (2021) A fast DST-Based Gerchberg-Saxton algorithm for Binary-Phase holography. *OSA Imaging and Applied Optics Congress 2021 (3D, COSI, DH, ISA, pcAOP)*, Jul., p. DW5E.6, Optical Society of America.

[27] Chakravarthula, P., Peng, Y., Kollin, J., Fuchs, H., and Heide, F. (2019) Wirtinger holography for near-eye displays. *ACM Trans. Graph.*, **38**, 1–13.

[28] Wang, Y., Dong, D., Christopher, P. J., Kadis, A., Mouthaan, R., Yang, F., and Wilkinson, T. D. (2020) Hardware implementations of computer-generated holography: a review. *Organ. Ethic.*, **59**, 102413.

[29] Yaraş, F., Kang, H., and Onural, L. (2010) State of the art in holographic displays: a survey. *J. Display Technol.*.

[30] Dong, D., Wang, Y., Kadis, A., and Wilkinson, T. D. (2020) Cost-optimized heterogeneous FPGA architecture for non-iterative hologram generation. *Appl. Opt.*, **59**, 7540–7546.

[31] Peng, Y., Choi, S., Padmanaban, N., and Wetzstein, G. (2020) Neural holography with camera-in-the-loop training. *ACM Trans. Graph.*, **39**, 1–14.

[32] Liu, S. and Takaki, Y. (2020) Optimization of Phase-Only Computer-Generated holograms based on the gradient descent method. *NATO Adv. Sci. Inst. Ser. E Appl. Sci.*, **10**, 4283.

[33] Chen, C., Lee, B., Li, N.-N., Chae, M., Wang, D., Wang, Q.-H., and Lee, B. (2021) Multi-depth hologram generation using stochastic gradient descent algorithm with complex loss function. *Opt. Express*, **29**, 15089–15103.

[34] Robbins, H. and Monro, S. (1951) A stochastic approximation method. *Ann. Math. Stat.*, **22**, 400–407.

[35] Paszke, A., et al. (2019) PyTorch: An imperative style, high-performance deep learning library. *Adv. Neural Inf. Process. Syst.*, **32**.

[36] Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., and Lerer, A. (2017) Automatic differentiation in PyTorch.

[37] Kadis, A., Wang, Y., Dong, D., Christopher, P., Mouthaan, R., and Wilkinson, T. D. (2021) HoloBlade: an open-hardware spatial light modulator driver platform for holographic displays. *Appl. Opt.*, **60**, A313–A322.

[38] Chang, K.-H., Seder, T., Hall, J., and Thompson, J. (2021) Speckle-reduced holographic projection with a piston-mode spatial light modulator. *Emerging Digital Micromirror Device Based Systems and Applications XIII*, Mar., vol. 11698, pp. 133–139, SPIE.

[39] Isomae, Y., Sugawara, N., Iwasaki, N., Ohkawa, S., Xiao, X., Honda, T., and Amari, K. (2022) Compact phase-only spatial light modulator with pixel pitch of 4.25 μm and high photostability. *Practical Holography XXXVI: Displays, Materials, and Applications*, Mar., vol. 12026, pp. 65–70, SPIE.

[40] Lee, B., Kim, D., Lee, S., Chen, C., and Lee, B. (2022) High-contrast, speckle-free, true 3D holography via binary CGH optimization. *Sci. Rep.*, **12**, 2811.

[41] Matsushima, K. and Shimobaba, T. (2009) Band-limited angular spectrum method for numerical simulation of free-space propagation in far and near fields. *Opt. Express*, **17**, 19662–19673.

[42] Kingma, D. P. and Ba, J. (2014) Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Author Biography

Andrew Kadis graduated from the University of Adelaide, Australia (2010), with a first-class degree in engineering and a bachelor of computer science. Before commencing his PhD studies in 2018, he had considerable experience in the industry, working on embedded systems in drones, medical devices, and life sciences equipment. He is currently at Cambridge University working with Professor Tim Wilkinson in the CMMPE group.