Data in Brief 54 (2024) 110424

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

A human erythrocytes hologram dataset for learning-based model training

Raul Castañeda[a], Carlos Trujillo[a], Ana Doblas[b,*]

[a] Applied Optics Group, School of Applied Sciences and Engineering EAFIT University, Medellin 050037, Colombia
[b] Electrical and Computer Engineering Department, University of Massachusetts – Dartmouth, USA

ARTICLE INFO

Article history:
Received 19 January 2024
Revised 9 April 2024
Accepted 9 April 2024
Available online 15 April 2024

Dataset link: human RBCs holograms (Original data)

Keywords:
Experimental recording
Digital holographic microscopy
Quantitative phase imaging
Interferometric techniques
Biological specimens
Red blood cells

ABSTRACT

This manuscript presents a paired dataset with experimental holograms and their corresponding reconstructed phase maps of human red blood cells (RBCs). The holographic images were recorded using an off-axis telecentric Digital Holographic Microscope (DHM). The imaging system consists of a 40 × /0.65NA infinity-corrected microscope objective (MO) lens and a tube lens (TL) with a focal distance of 200 mm, recording diffraction-limited holograms. A CMOS camera with dimensions of 1920 × 1200 pixels and a pixel pitch of 5.86 µm was located at the back focal plane of the TL lens, capturing image-plane holograms. The off-axis, telecentric, and diffraction-limited DHM system guarantees accurate quantitative phase maps. Initially comprising 300 holograms, the dataset was augmented to 36,864 instances, enabling the investigation (i.e., training and testing) of learning-based models to reconstruct aberration-free phase images from raw holograms. This dataset facilitates the training and testing of end-to-end models for quantitative phase imaging using DHM systems operating at the telecentric regime and non-telecentric DHM systems where the spherical wavefront has been compensated physically. In other words, this dataset holds promise for advancing investigations in digital holographic microscopy and computational imaging.

* Corresponding author.
E-mail address: adoblas@umassd.edu (A. Doblas).

https://doi.org/10.1016/j.dib.2024.110424

Specifications Table

Subject	*Optics*
Specific subject area	Digital holographic microscopy (DHM) DHM is a cutting-edge imaging method that merges holography and microscopy to reconstruct high-resolution 3D images of samples in life and material sciences.
Data format	Raw, Reconstructed, Filtered.
Type of data	Recorded digital holograms, PNG format. Reconstructed phase images, PNG format.
Data collection	Data collection was performed using an off-axis Mach-Zehnder interferometer whose light source was a diode laser with a central wavelength equal to 532 nm. The imaging system in the DHM setup was composed of a 40 $\times$ /0.65NA infinity-corrected microscope objective lens and a tube lens with a focal length of 200 mm. The holograms were captured by a CMOS camera, featuring 1920 $\times$ 1200 square pixels with a side length of 5.86 µm. The off-axis configuration enables the reconstruction of both amplitude and phase images using a single hologram. Quantitative phase images were computationally obtained by applying an automatic method that finds the best-reconstructed phase images, which corresponds to the one with the minimum phase distortions [1].
Data source location	**Institution:** University of Memphis **City/Town/Regions:** Memphis, TN. **Country:** USA
Data accessibility	Repository name: The Open Science Framework (OSF) Data identification number: doi: 10.17605/OSF.IO/8P7BA Direct URL to data: https://osf.io/8p7ba/?view_only=e8d752fda65e456f896b72a91aa8db51
Related research article	[1] R. Castaneda, C. Trujillo, and A. Doblas, “Video-rate quantitative phase imaging using a digital holographic microscope and a generative adversarial network,” Sensors 21, 1–16 (2021).

1. Value of the Data

- This dataset benefits researchers in the field of digital holographic microscopy (DHM), offering a rich experimental dataset for developing, training, validating, and testing new learning-based models. The availability of such a dataset is particularly advantageous for developing end-to-end deep-learning research models, providing a robust foundation to advance learning-based models for real-time quantitative phase imaging from raw holograms with the need for pre- or post-numerical procedures.
- The proposed dataset is the sole available dataset in the literature for training end-to-end deep learning models in quantitative phase imaging using an off-axis DHM system operating at the telecentric regime and a non-telecentric regime with a physical compensation approach to remove the spherical wavefront generated by the optical configuration between the microscope objective lens and the tube lens.
- Transfer learning methods can be applied to pre-trained deep learning models that reconstruct quantitative phase images of red blood cells from this dataset to generate quantitative phase images of different specimens.
- The dataset, which consists of pairs of holograms and reconstructed phase images, promotes the reproducibility of any research study by offering a freely available dataset for training various models, guaranteeing the replicability of experiments and outcomes.
- This dataset holds potential for future integration with existing databases to aid in discerning between healthy and unhealthy erythrocytes, potentially enhancing diagnostic capabilities for blood-related disorders in clinical settings.

- Future research directions could be focused on (1) expanding the database of holograms for training deep learning models on quantitative phase imaging; (2) exploring advanced techniques for data augmentation in DHM to enhance the diversity and richness of the dataset, facilitating more robust model training; (3) investigating different deep learning architectures tailored specifically for DHM, which may lead to improved accuracy in phase reconstructions; and (4) physical integration of real-time holographic data acquisition and model deployment, enabling applications in dynamic environments.

2. Background

Digital holographic microscopy (DHM) allows us to capture and reconstruct the full complex field information of microscopic samples, turning DHM into a phase retrieval instrument for quantitative phase imaging (QPI) [2–7]. Over the last few years, significant advances in Deep Learning strategies have unleashed the synergy of these models to improve the potential of DHM applications by providing real-time reconstructed amplitude and phase images from raw holograms [8–14]. In 2021, we proposed a conditional generative adversarial network (cGAN) to reconstruct quantitative free-of-aberration phase images from biological samples using a transmission off-axis DHM [1]. The cGAN model outperforms the conventional computational DHM reconstruction method [15], generating quantitative phase images with stable background levels and minimum phase distortions regardless of the number of cells in the imaged area and the experimental conditions.

None of the reported learning-based models in digital holography [1,8-14] provide their dataset without restriction. This lack of open-access experimental datasets challenges the progress of learning-based models in holographic systems, requiring the generation of datasets before investigating any learning-based model. This work provides the first experimental open-access DHM dataset.

3. Data Description

The dataset contains 36,864 pairs of holograms and their respective reconstructed phase images, meaning a total of 73,728 images. The available dataset (holograms and phase images) is stockpiled in a .zip file available for free download from DOI 10.17605/OSF.IO/8P7BA. The database is organized in two main folders: 'Holograms' and 'Phase'. Each folder contains two subfolders containing the respective images for 'Training' and 'Testing', allocating 80% for training and the remaining 20% for testing. This means the "Training" subfolder has 29,491 images and the "Testing" one has 7,373 images, ensuring a balanced distribution for robust model development. Users are encouraged to adjust the distribution within each subfolder based on their model training design. Fig. 1 illustrates the dataset's structure and examples of the images, showcasing holograms, and their corresponding phase maps.

4. Experimental Design, Materials and Methods

The dataset was generated through a two-stage process. The first stage involved recording off-axis holograms using an experimental Mach-Zehnder interferometer operating in transmission mode, as illustrated in Fig. 2. This instrument represents the fundamental implementation of DHM, a technique with diverse applications across various fields [2–7]. Within this configuration, a 532-nm laser serves as the illumination source. The collimated emerging beam from the laser is expanded by an afocal 4f system. The expanded illuminating wavefront is divided into two wavefronts (i.e., reference and object wavefronts) using a cube beam splitter (BS1). The

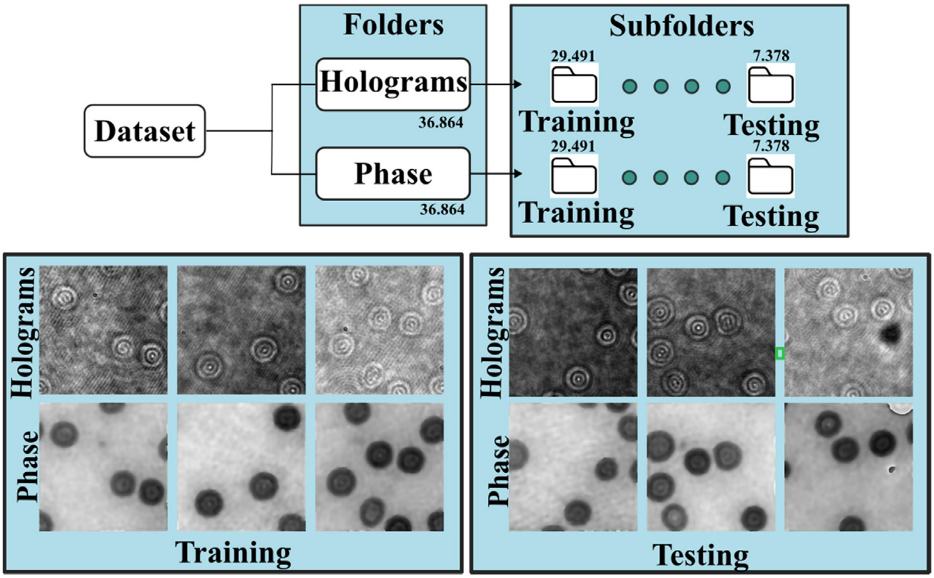

Fig. 1. Dataset structure.

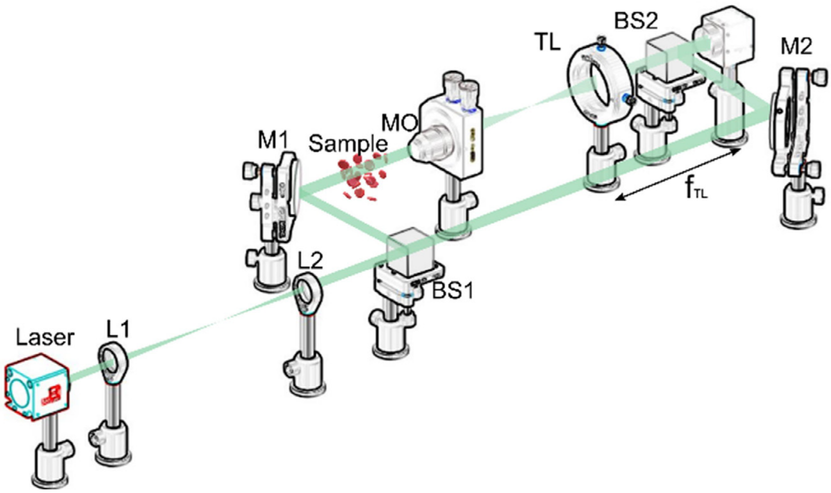

Fig. 2. Mach-Zehnder-based digital holographic microscopy setup.

object wavefront illuminates the sample, and the light scattered by it passes through an imaging system comprising a 40 × /0.65NA infinity-corrected microscope objective (MO) lens and a tube lens (TL) with a focal distance of 200 mm. The sample and detector planes are conjugated, providing an image-plane configuration (i.e., in-focus images are recorded). To guarantee reconstructed phase images free of spherical aberration related to the optical configuration between the MO and TL lenses in the DHM system, the imaging system between the MO and TL lenses operates in the telecentric regime (i.e., the back focal plane of the MO lens coincides with the front focal plane of the TL lens) [16–17]. Telecentric DHM systems are linear shift-invariant [16], meaning that the system's response is independent of the sample position as opposed to

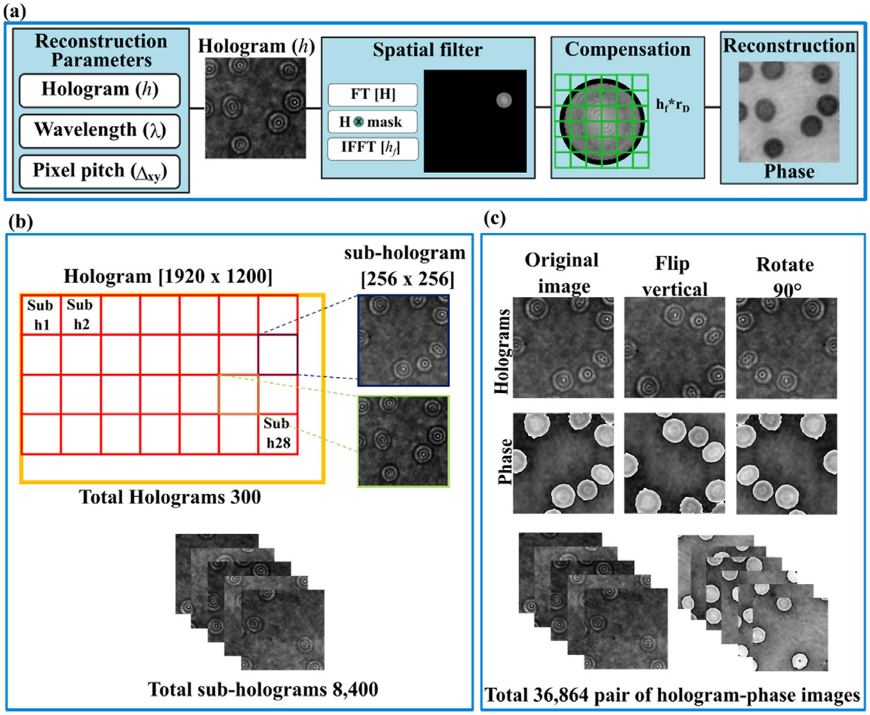

Fig. 3. (a) Computational framework of the Trujillo's method [15]. (b) Proposed framework to generate our augmented dataset from 300 paired images of 1920 × 1200-px^2 to 8,400 paired images of 256 × 256 px^2. (c) Illustration of the five image transformations applied to the data augmentation procedure.

non-telecentric DHM systems. In other words, accurate quantitative phase maps are intrinsically achieved if the DHM system operates in the telecentric regime. The advantage of telecentric versus non-telecentric DHM systems in quantitative phase imaging of biological samples was shown in Ref. [17], demonstrating that telecentric DHM systems provide quantitative phase maps with a larger field of view and a smaller number of computational operations. In addition, the telecentric DHM system operates at the diffraction limit (i.e., no overlapping between the different diffraction terms in the hologram's spectrum), ensuring that the resolution capability of the DHM system coincides with the diffraction limit [18].

On the other hand, the reference wavefront propagates without perturbation to a second beam splitter (BS2) that recombines both the object and reference wavefronts. The superposition of both wavefronts generates an interference pattern, called the hologram, which is recorded using a digital sensor. The sensor was a CMOS camera of 1920 × 1200 pixels and 5.86-µm pixel pitch for our experiment. The DHM system operates in an off-axis architecture, providing reconstructed phase images from a single hologram. The off-axis architecture can be achieved by tilting the BS2 and the mirror (M2) in the DHM setup. To attain diffraction-limited resolution and image quality, a precise inclination of these elements is set to meet the diffraction-limited regime [18].

We have adopted the computational method proposed by Trujillo et al. [15] to reconstruct the phase images from the recorded holograms. This method automatically finds the best reconstructed phase image for off-axis holograms operating in the telecentric regime. Fig. 3(a) shows the schematic of Trujillo's computational method. The inputs for the method are the holograms, the wavelength of the illuminating laser source (λ), and the pixel pitch of the digital sensor (Δ_{xy}). In summary, Trujillo's method follows four key steps:

(i) The hologram undergoes Fourier transformation and subsequent spatial filtering to select the +1 diffraction term, eliminating both the DC and -1 terms.
(ii) The method identifies the spatial frequency positions (i.e., u_{max}, v_{max}) of the maximum peak in the spatially-filtered +1 diffraction term of step i).
(iii) The method initiates a square grid search to assess proper phase reconstruction, determining the optimal pair of values for (u'_{max}, v'_{max}) through a summation and thresholding metric applied to the phase maps reconstructed at every spatial frequency pair within the grid.
(iv) Finally, the method outputs the best phase images based on the maximum value of the summation and thresholding metric. The complete details and full Python implementation of the numerical processing to properly reconstruct off-axis holograms are found in Ref. [19]. The pyDHM library also provides two new methods to reconstruct automatic phase images free of aberrations.

In summary, the key features of the experimental holograms are: (1) the holograms have been acquired in an off-axis configuration, meaning that the different terms in the hologram's spectrum do not overlap and, therefore, quantitative phase maps can be reconstructed using a single hologram [20] as opposed to in-line or slightly off-axis DHM systems which required at least 3 or 2 holograms, respectively; (2) the DHM system operates in telecentric regime, providing intrinsically the same system's response regardless of the sample position within the field of view (i.e., a linear shift-invariant system) [16]. As a consequence of the telecentric configuration, the reconstruction framework in telecentric DHM system is simplified, not requiring the compensation of any additional spherical wavefront related to the optical configuration, and the hologram does not have to exhibit an area free of specimen to retrieve the parameters of any residual spherical wavefront [17]; and (3) the DHM system operates at diffraction limit, achieving the same resolution capability as the native coherent microscope [18].

Initially, 300 holograms were recorded using the previously described DHM system and then reconstructed using Trujillo's method [15] to obtain their quantitative phase images without or minimum phase distortions, see Fig. 3(a). The 1920 $\times$ 1200 px^2 paired images (hologram and reconstructed phase images) were systematically cropped into patches of 256 $\times$ 256 px^2 size, resulting in an expanded dataset of 8,400 sub-images (sub-holograms and phase images). Borders were removed from the images in each dimension to enhance diversity and robustness after the cropping process, as illustrated in Fig. 3(b). Additionally, data augmentation was implemented, involving random geometric transformations such as 90 or 180-degree image rotations and horizontal or vertical flipping, as illustrated in Fig. 3(b). Since five data augmentation transformations were applied to the image pairs, an initial count of 42,000 pairs of hologram-phase images was expected. However, a subsequent refinement process was required to remove images that did not contain any red blood cells or images with inadequate phase compensation, resulting in 36,864 pairs of hologram-phase images. The augmentation solely alters the direction of the holograms, without affecting their shape or the distribution of information within them. Therefore, the fidelity of data in the augmented holograms remains intact, ensuring the reliable performance of the models trained on the proposed dataset.

The hallmark of DHM versus traditional microscopy is that DHM allows us to reconstruct the three-dimensional topography of unstained samples from a single recorded hologram. In other words, the reconstructed phase image provides information on the specimen's morphology, topography, and chemical composition via the refractive index. The phase information (φ) is related to the axial topography (t) and the refractive index of the sample (n_s) via $\varphi = 2\pi\ (n_s - n_{im})\ t\ /\ \lambda$ where λ is the wavelength of the illuminating DHM source and n_{im} is the refractive index of sample surrounding media. Such information is not accessible to traditional bright-field microscopy.

In 2021, we used this dataset for training and testing a learning-based model based on a conditional Generative Adversarial Networks (cGAN) architecture [1]. That work demonstrated a cGAN model as a video-rate quantitative phase imaging tool to reconstruct quantitative phase images from raw holograms. The reconstructed phase images generated by the trained cGAN model have stable background levels and minimum phase distortions without needing pre- or post-numerical procedures. In that work, the trained model using the augmented dataset was validated on a second dataset with 7,005 paired hologram/phase images. Although this second dataset was also recorded using an off-axis DHM system operating in the telecentric regime, the experimental conditions (i.e., interference angle and intensity in the object and reference beams) differed from the training dataset. Our results in Ref. [1] show the effectiveness of the trained cGAN model for off-axis telecentric holograms at different experimental conditions, confirming that the original dataset is not biased through data augmentation. Furthermore, in that work, we demonstrate that (1) the cGAN model provides quantitative phase images regarding the number of cells in the imaged area) and (2) the cGAN model outperforms the conventional DHM reconstruction framework.

Limitations

The limitations of this dataset are primarily related to constraints encountered during the acquisition process. For example, the cell density within the dataset was restricted to 9 cells within the field of view, potentially limiting the diversity of cellular configurations present in the imaged area. The applicability of the dataset to other holographic setups is restricted since our dataset was focused on off-axis DHM systems operating in the telecentric regime. Robust maintenance of sample conditions posed challenges experimentally, introducing undesired variability into the dataset. For example, temporal fluctuations of the DHM system may generate distorted phase images. Furthermore, while the dataset's size is substantial, it may not encompass the full range of possible variations due to constraints in sample preparation and holographic recording conditions. Finally, despite meticulous curation, subtle biases within the dataset arising from variations in sample characteristics cannot be entirely ruled out. These limitations highlight the inherent complexities in capturing holographic data of biological specimens, underscoring the ongoing need for advancements in data collection techniques to address these challenges effectively.

Ethics Statement

The authors affirm that the dataset presented in this work did not entail experiments on human subjects or involve animal experimentation. The samples were obtained through procedures that posed no risk to the individuals from whom the samples were derived. All protocols adhered to ethical guidelines, and necessary permissions were obtained from relevant ethical committees, ensuring compliance with ethical standards and requirements.

CRediT Author Statement

Raul Castaneda: Data acquisition, Hardware, Writing - Original Draft, Methodology; **Carlos Trujillo:** Supervision, Data curation, Writing - Review & Editing, Formal analysis, Methodology; **Ana Doblas:** Data acquisition, Hardware, Supervision, Investigation, Methodology, Writing - Review & Editing.

Data Availability

human RBCs holograms (Original data) (The Open Science Framework (OSF))

Acknowledgements

This research was funded by the National Science Foundation (NSF CAREER grant 2042563). A. Doblas acknowledges the support from her former (University of Memphis) and new (University of Massachusetts Dartmouth) institutions.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] R. Castaneda, C. Trujillo, A. Doblas, Video-rate quantitative phase imaging using a digital holographic microscope and a generative adversarial network, Sensors 21 (2021) 1–16.
[2] M.K. Kim, Digital holographic microscopy, Principles, Techniques, and Aplications, Springer, 2011.
[3] F.C. Cheong, B.J. Krishnatreya, D.G. Grier, Strategies for three-dimensional particle tracking with holographic video microscopy, Opt. Express 18 (2010) 13563–13573.
[4] F. Dubois, C. Yourassowsky, O. Monnom, J.-C. Legros, O. Debeir, P. Van Ham, R. Kiss, C. Decaestecker, Digital holographic microscopy for the three-dimensional dynamic analysis of in vitro cancer cell migration, J. Biomed. Opt. 11 (2006) 054032.
[5] B. Kemper, D. Carl, J. Schnekenburger, I. Bredebusch, M. Schäfer, W. Domschke, G. von Bally, Investigation of living pancreas tumor cells by digital holographic microscopy, J. Biomed. Opt. 11 (2006) 034005.
[6] M.K. Kim, Applications of digital holography in biomedical microscopy, J. Opt. Soc. Korea 14 (2010) 77–89.
[7] Y.K. Park, C. Depeursinge, G. Popescu, Quantitative phase imaging in biomedicine, Nat. Photonics 12 (2018) 578–589.
[8] Y. Rivenson, Y. Zhang, H. Günaydın, D. Teng, A. Ozcan, Phase recovery and holographic image reconstruction using deep learning in neural networks, Light. Sci. Appl. 7 (2) (2018) 17141.
[9] K. Wang, J. Dou, Q. Kemao, J. Di, J. Zhao, Y-Net: a one-to-two deep learning framework for digital holographic reconstruction, Opt. Lett. 44 (19) (2019) 4765–4768.
[10] I. Moon, K. Jaferzadeh, Y. Kim, B. Javidi, Noise-free quantitative phase imaging in Gabor holography with conditional generative adversarial network, Opt. Express 28 (18) (2020) 26284–26301.
[11] N. Chen, C. Wang, and W. Heidrich, "Holographic 3D particle imaging with model-based deep network," in IEEE Transactions on Computational Imaging, vol. 7, pp. 288-296, 2021.
[12] J. Di, J. Wu, K. Wang, J. Tang, Y. Li, J. Zhao, Quantitative phase imaging using deep learning-based holographic microscope, Front. Phys. 9 (2021) 1–7.
[13] Wang W, Ali N, Ma Y, Dong Z, Zuo C, and Gao P, "Deep learning-based quantitative phase microscopy," Front. Phys. 11:1218147.
[14] S. Ma, Q. Liu, Y. Yu, Y. Luo, S. Wang, Quantitative phase imaging in digital holographic microscopy based on image inpainting using a two-stage generative adversarial network, Opt. Express 29 (16) (2021) 24928–24946.
[15] C. Trujillo, R. Castañeda, P. Piedrahita-Quintero, J. Garcia-Sucerquia, Automatic full compensation of quantitative phase imaging in off-axis digital holographic microscopy, Appl. Opt. 55 (2016).
[16] A. Doblas, E. Sánchez-Ortiga, M. Martínez-Corral, G. Saavedra, P. Andrés, J. Garcia-Sucerquia, Shift-variant digital holographic microscopy: inaccuracies in quantitative phase imaging, Opt. Lett. 38 (2013) 1352–1354.
[17] A. Doblas, E. Sánchez-Ortiga, M. Martínez-Corral, G. Saavedra, J. Garcia-Sucerquia, Accurate single-shot quantitative phase imaging of biological specimens with telecentric digital holographic microscopy, J. Biomed. Opt. 19 (2014) 46022.
[18] E. Sánchez-Ortiga, A. Doblas, G. Saavedra, M. Martínez-Corral, J. Garcia-Sucerquia, G. Saavedra, J. Garcia-Sucerquia, Off-axis digital holographic microscopy: practical design parameters for operating at diffraction limit, Appl. Opt. 53 (2014) 2058–2066.
[19] R. Castañeda, C. Trujillo, A. Doblas, pyDHM: a python library for applications in digital holographic microscopy, PLoS One 17 (2022) 1–23.
[20] E. Cuche, P. Marquet, C. Depeursinge, Spatial filtering for zero-order and twin-image elimination in digital off-axis holography, Appl. Opt. 39 (2000) 4070–4075.