

# REMOVAL OF THE TWIN IMAGE ARTIFACT IN HOLOGRAPHIC LENS-FREE IMAGING BY SPARSE DICTIONARY LEARNING AND CODING

Benjamin D. Haefele<sup>†</sup>, Sophie Roth<sup>‡</sup>, Lin Zhou<sup>‡</sup>, and René Vidal<sup>†</sup>

<sup>†</sup>Center for Imaging Science, Johns Hopkins University. Baltimore, MD, USA.

<sup>‡</sup>Imec. Leuven, Belgium.

## ABSTRACT

Mitigating the effects of the twin image artifact is one of the key challenges in holographic lens-free microscopy. This artifact arises due to the fact that imaging detectors can only record the magnitude of the hologram wavefront but not the phase. Prior work addresses this problem by attempting to simultaneously estimate the missing phase and reconstruct an image of the object specimen. Here we explore a fundamentally different approach based on post-processing the reconstructed image using sparse dictionary learning and coding techniques originally developed for processing conventional images. First, a dictionary of atoms representing characteristics from either the true image of the specimen or the twin image are learned from a collection of patches of the observed images. Then, by expressing each patch of the observed image as a sparse linear combination of the dictionary atoms, the observed image is decomposed into a component that corresponds to the true image and another one that corresponds to the twin image artifact. Experiments on counting red blood cells demonstrate the effectiveness of the proposed approach.

**Index Terms**— Holography, lens-free imaging, sparse coding, dictionary learning, blood cell counting

## 1. INTRODUCTION

Holographic lens-free imaging (LFI) is a technique that reconstructs images of a specimen from holographic diffraction patterns that are generated by passing a coherent light source (e.g., a laser) through the specimen. In microscopy applications, LFI has several advantages over conventional techniques. First, because there are no lenses in the imaging system, its overall cost and physical size can be greatly reduced compared to traditional microscopes. Second, LFI allows much wider fields of view to be imaged than a conventional microscope with equal magnification. Third, because the image of the specimen is generated through post-processing the recorded diffraction pattern, there is no need for an operator to manually focus the system as the focal depth can be adjusted automatically through post-processing.

These advantages have led to LFI being explored as a potential method to track or detect various types of cells in so-

lution over a large field of view (and potentially in 3 dimensions) with notable success [1]. However, many of these techniques either rely on complex models of the light diffraction process, which can be expensive to fit computationally, or assume that the cells to be detected/tracked are sufficiently separated so that the holograms generated by the cells have minimal interaction and can be easily fit assuming the holograms are generated independently of other cells in the volume [1].

When the concentration of cells (or more generally objects) increases, the holograms generated by different objects begin to combine before being detected by the imaging sensor. While the superposition of diffraction patterns could be resolved if imaging sensors recorded both the magnitude and phase of the hologram wavefront, usually only the magnitude is recorded. As a result, images reconstructed via traditional holographic image reconstruction techniques that do not attempt to estimate the missing phase are typically corrupted by significant artifacts that are a consequence of not capturing the phase information of the diffraction pattern. This problem is commonly referred to as the *twin image* artifact, which typically manifests as wave-like distortions emanating from the true object (see the top row of Fig. 1 for an example).

To address this issue, prior work has focused largely on techniques to estimate the missing phase information followed by traditional holographic reconstruction techniques. However, the problem of estimating both the reconstructed image along with the missing phase is typically an ill-posed and under-determined inverse problem, hence assumptions regarding the expected statistical or geometrical properties of the reconstructed image are typically necessary to bound the space of solutions and make the problem tractable. A common assumption that is well suited to many applications in holographic imaging is to assume that the true image of the specimen is *sparse* (i.e., most of the pixels in the true image do not contain objects) [2, 3]. Here we explore a much simpler method which likewise operates under the assumption that the true image of the object is sparse but takes a fundamentally different approach from phase recovery techniques. In particular, instead of attempting to recover the missing phase as a part of the reconstruction, we seek to remove the twin image artifact through a post-processing step based on *sparse dictionary learning and coding*, which allows one to

separate a reconstructed image into components corresponding largely to the true image of the object and the twin image artifact in an unsupervised manner. We demonstrate the effectiveness of our approach in counting the concentration of red blood cells in samples of human anti-coagulated blood.

## 2. METHODS

This section describes the proposed approach to minimizing the effects of the twin image artifact in images obtained using traditional holographic reconstruction methods. At a high level our approach consists of three main steps. First, we use sparse dictionary learning techniques to learn a suitable representation for the images in our dataset. Then we automatically separate our learned dictionary into elements corresponding to either the true image or the twin image. Finally we use this learned and separated dictionary to decompose new images into two components: one containing the true image and the other one containing the twin image artifact. As our experiments will show, this decomposition allows one to accurately count the number of red blood cells in a holographic image via a simple thresholding approach applied to the true image.

### 2.1. Holographic Image Collection and Reconstruction

The proposed approach will be evaluated on images reconstructed from 296 diffraction holograms of anti-coagulated human blood from 5 healthy volunteers ( $\sim 60$  images per donor) diluted at a 300:1 ratio suspended in a micro-fluidic channel using a lens-free imaging system [4]. The images were reconstructed using the wide-angular spectrum approximation for the image diffraction process [5]. Specifically, if  $H$  is the recorded hologram (the square-root of the raw pixel values since the CCD records the squared amplitude of the wavefront), the image  $I$  was reconstructed as  $I = |T(z) * H|$  where  $*$  is the 2D convolution operator and  $T(z)$  is the wide-angular spectrum transfer function at a focal depth  $z$  [5]. The top row of Fig. 1 shows sample reconstructed images. Note that cells (predominately red blood cells) are clearly visible as dark objects in the image surrounded by the wave-like twin image artifact. To minimize the effects of these artifacts we will employ a sparse dictionary learning and coding method, as described next.

### 2.2. Sparse Dictionary Learning and Coding

Sparse dictionary learning [6] is a well established technique for many applications in image processing and is based on modeling small patches (e.g., 20x20 pixels) extracted from an image as a linear combination of elements (also referred to as atoms or components) from a dictionary. The total number of elements in the dictionary can potentially be very large, for example larger than the dimensionality of the extracted

patches, in which case the dictionary is over-complete. Therefore, the model also seeks to find *sparse* solutions, which limit the number of dictionary elements used to represent any given patch (or the number of dictionary elements used is sparse). Specifically, if we are given a suitable dictionary,  $D \in \mathbb{R}^{m \times r}$ , where  $m$  is the patch dimension and  $r$  is the number of dictionary atoms, then the goal of the sparse coding model<sup>1</sup> is to solve a problem of the form

$$\min_{\alpha} \frac{1}{2} \|x - D\alpha\|_2^2 + \gamma \|\alpha\|_1, \quad (1)$$

where  $x \in \mathbb{R}^m$  is a patch extracted from the image, the  $\ell_1$  norm is used as a regularization function to encourage sparse solutions, and  $\gamma$  is a non-negative scalar that balances the trade-off between fitting the data and promoting sparse solutions [6]. The above formulation is convex in  $\alpha$  and easily solved by many efficient algorithms [7, 8]. Since the dictionary,  $D$ , is typically not known *a priori*, the *dictionary learning* problem takes a collection of  $N$  patches,  $X = [x_1, \dots, x_N] \in \mathbb{R}^{m \times N}$ , extracted from an image (or a collection of images) and seeks to solve an optimization problem jointly over both the encoding variables,  $A = [\alpha_1, \dots, \alpha_N] \in \mathbb{R}^{r \times N}$ , and the dictionary,  $D \in \mathbb{R}^{m \times r}$ , of the form

$$\min_{D, A} \frac{1}{2} \|X - DA\|_F^2 + \gamma \|A\|_1 \text{ s.t. } \|D_i\|_2 = 1 \quad \forall i \in [1, r]. \quad (2)$$

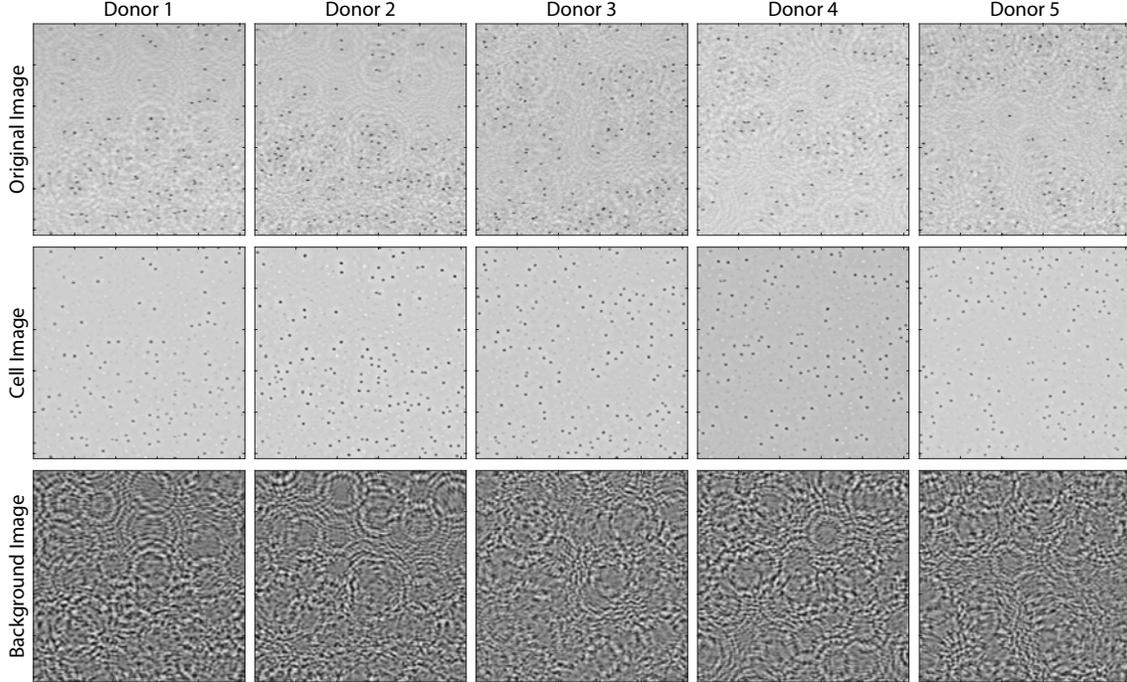
We applied this general dictionary learning framework to images reconstructed from diffraction holograms using traditional holographic reconstruction techniques described in §2.1. To learn the dictionary, we extracted all possible patches of size 20x20 pixels (which are larger than the typical size of blood cells) from a 512x512 crop from a single image using a sliding window with a stride of 1. The patches were then normalized to have zero mean and unit  $\ell_2$  norm. The dictionary was then learned using the publicly available SPAMS<sup>2</sup> software package [9] with the parameter  $\gamma$  set to 0.15 in (2). Fig. 2 shows the result from learning a dictionary with 625 atoms. Note that many of the learned dictionary atoms correspond to cells (approximately top 5 rows), while the rest correspond to the twin image artifact. Note that the dictionary shown in Fig. 2 was automatically sorted to identify atoms that correspond to cells versus background as described next.

### 2.3. Dictionary Separation

Once the dictionary has been learned, we automatically separate it into atoms that correspond to “cells” (since the images are of human blood) and atoms that correspond to the “background”, which are largely due to the twin image artifact. Specifically, by taking patches of size 20x20, cell atoms

<sup>1</sup>Note that there are many potential sparse coding models possible, but (1) describes the well-known Lasso or Matching Pursuit model.

<sup>2</sup><http://spams-devel.gforge.inria.fr/>



**Fig. 1.** Example image decompositions for one example image from each of the five blood donors. (Top Row) Original images. (Middle Row) Images reconstructed from cell dictionary atoms and coefficients. (Bottom Row) Images reconstructed from background dictionary atoms and coefficients. Note that the images have different gray-scale scalings to improve contrast.

only contain a small portion of the patch which is significantly different from the background intensity, whereas background atoms are characterized by wave-like components at various orientations typical of the twin image artifact, which are largely different from the background intensity at every pixel. This suggests that cell dictionary atoms are sparser than background dictionary atoms, and simply sorting the dictionary atoms based on their  $\ell_1$  norms produces a robust separation between cell atoms and background atoms as shown in Fig. 2. The final segmentation of the dictionary atoms into cell and background atoms was made by choosing atoms with a  $\ell_1$  norm below a manually chosen threshold as cell atoms.

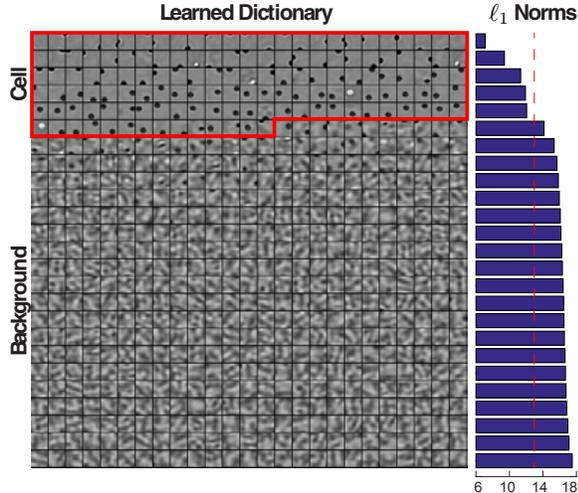
#### 2.4. Image Decomposition

Once the learned dictionary has been separated into cell and background components, new images are processed by encoding  $20 \times 20$  patches extracted from the image via the sparse coding formulation given in (1). Then the reconstructed image is approximated as a decomposition into cell and background component images using the corresponding dictionary elements and sparse coefficients. Specifically, given a sorted dictionary,  $D = [D_{cell} \ D_{back}]$ , the problem in (1) is solved for all the patches in an image, producing sparse coefficients which likewise correspond to cell and background components,  $A = [A_{cell}^T \ A_{back}^T]^T$ . From this representation, the cell and background components of the image are reconstructed from the patch decompositions  $D_{cell}A_{cell}$  and  $D_{back}A_{back}$ ,

respectively. Examples of this decomposition are shown in the second and third rows of Fig. 1. These images were created by extracting and encoding all possible patches in the original image using a sliding window with a stride of 1 and then reconstructing the images by returning the patches to their original locations and taking the mean of the pixels across patches where they overlapped. To reconstruct images for estimating red blood cell concentrations the patches were extracted using a non-overlapping sliding window with a stride of 20 to improve the computational efficiency of the method. Note that in general the twin image background artifact is largely removed from the cell images. The most prominent artifacts still remaining in the cell image are areas of the twin image that happen to be largely circular and hence can be efficiently represented by the circular cell dictionary atoms.

### 3. EXPERIMENTAL DETAILS AND RESULTS

To evaluate the effectiveness of our approach on a medically significant task, we explored the use of our image decomposition algorithm in the estimation of the concentration of red blood cells from lens-free holographic images. After applying the image decomposition algorithm described in the previous section, we estimated the number of blood cells present in a given image by thresholding the cell component image and counting the number of particles in the thresholded image with an area larger than a given size. Using the estimated number of cells in a given image, the red blood cell



**Fig. 2.** (Left) Learned dictionary sorted by  $\ell_1$  norm (in ascending order). The first 139 atoms were used as the cell dictionary (outlined in red) and the remaining 486 atoms as the background dictionary. (Right)  $\ell_1$  norms of the last dictionary atom in each row. The cutoff threshold for the cell/background separation was 13 (red-dashed line).

concentration estimate for a given image was calculated from the known volume of the micro-fluidic channel and dilution factor. We estimated the red blood cell concentration for a particular blood donor by taking the median of the estimated red blood cell concentrations over approximately 60 images of blood collected from the donor (the number of images per donor in the dataset ranged from 58-62). Processing the images with non-overlapping patches took approximately 1.8 seconds per image on a laptop with a 2.8GHz i7-3840QM processor using an image crop size of 1280x4000 pixels. To establish a baseline for our image decomposition algorithm we also estimated red blood cell concentrations by thresholding the original reconstructed images after normalizing the images to have unit standard deviation. In both cases we chose the value of the threshold and minimum particle size via leave-one-donor-out cross validation by comparing red blood cell concentrations estimated from the lens-free images to red blood cell concentrations measured via a laboratory hematology analyzer. Additionally, for the original image we also report results for keeping the minimum particle size fixed at 0 and only cross validating over the threshold. The cross validation errors for each of the five donors are shown in Table 1. Note that our proposed method significantly improves the accuracy and reliability of estimating red blood cell concentration over the original reconstructed image.

#### 4. CONCLUSIONS

We have presented a method to remove twin image artifacts from reconstructed holographic lens-free images based on sparse dictionary learning and coding and demonstrated the

advantages of our approach in estimating the number of red blood cells present in images of human blood. While in this paper we applied our method to images of red blood cells, we anticipate it will be applicable to any holographic image that consists of multiple small objects with similar appearances.

**Acknowledgements.** The authors thank Richard Stahl, Geert Vanmeerbeeck, Murali Jayapala and Stuart Ray for insightful discussions. This work was funded by miDIAGNOSTICS.

#### 5. REFERENCES

- [1] Pasquale Memmolo, Lisa Miccio, Melania Paturzo, Giuseppe Di Caprio, Giuseppe Coppola, Paolo A. Netti, and Pietro Ferraro, "Recent advances in holographic 3d particle tracking," *Advances in Optics and Photonics*, vol. 7, no. 4, pp. 713–755, 2015. 1
- [2] Jun Song, Christine Leon Swisher, Hyungsoon Im, Sangmoo Jeong, Divya Pathania, Yoshiko Iwamoto, Misha Pivovarov, Ralph Weissleder, and Hakho Lee, "Sparsity-based pixel super resolution for lens-free digital in-line holography," *Scientific reports*, vol. 6, 2016. 1
- [3] Loïc Denis, Dirk Lorenz, Eric Thiébaud, Corinne Fournier, and Dennis Trede, "Inline hologram reconstruction with sparsity constraints," *Optics letters*, vol. 34, no. 22, pp. 3475–3477, 2009. 1
- [4] Richard Stahl, Geert Vanmeerbeeck, Gauthier Lafruit, Roeland Huys, Veerle Reumers, Andy Lambrechts, Chao-Kang Liao, Chin-Chun Hsiao, Masayuki Yashiro, Masashi Takemoto, et al., "Lens-free digital in-line holographic imaging for wide field-of-view, high-resolution and real-time monitoring of complex microscopic objects," in *SPIE BiOS*. International Society for Optics and Photonics, 2014, pp. 89471F–89471F. 2
- [5] Myung K Kim, *Digital Holographic Microscopy*, Springer, 2011. 2
- [6] Michael Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, Springer, 2010. 2
- [7] Bradley Efron, Trevor Hastie, Iain Johnstone, Robert Tibshirani, et al., "Least angle regression," *The Annals of statistics*, vol. 32, no. 2, pp. 407–499, 2004. 2
- [8] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004. 2
- [9] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online dictionary learning for sparse coding," in *International Conference on Machine Learning*, 2009. 2

**Table 1.** Cross validation errors in red blood cell concentration estimation using thresholding of original and cell images.

Donor	Original Im.	Org. Im. (min size=0)	Cell Im.
1	-41.4%	-41.4%	1.9%
2	-8.7%	-5.6%	4.1%
3	-24.1%	-24.1%	-5.6%
4	-31.0%	-31.0%	-0.8%
5	81.0%	13.0%	1.4%
Mean  Error	37.3 %	23.0%	2.8%