# Random-subset fitting of digital holograms for fast three-dimensional particle tracking

The Harvard community has made this article openly available. **Please share** how this access benefits you. Your story matters

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Published Version</td>
<td>doi:10.1364/AO.53.00G177</td>
</tr>
<tr>
<td>Citable link</td>
<td><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:12748673">http://nrs.harvard.edu/urn-3:HUL.InstRepos:12748673</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a></td>
</tr>
</tbody>
</table>
Random-subset fitting of digital holograms for fast three-dimensional particle tracking [Invited]

Thomas G. Dimiduk,1 Rebecca W. Perry,2 Jerome Fung,1 and Vinothan N. Manoharan2,1,*

1Department of Physics, Harvard University, 17 Oxford St., Cambridge, Massachusetts 02138, USA
2Harvard School of Engineering and Applied Sciences, 29 Oxford St., Cambridge, Massachusetts 02138, USA
*Corresponding author: vnm@seas.harvard.edu

Received 28 April 2014; revised 19 July 2014; accepted 19 July 2014; posted 21 July 2014 (Doc. ID 211015); published 27 August 2014

Fitting scattering solutions to time series of digital holograms is a precise way to measure three-dimensional dynamics of microscale objects such as colloidal particles. However, this inverse-problem approach is computationally expensive. We show that the computational time can be reduced by an order of magnitude or more by fitting to a random subset of the pixels in a hologram. We demonstrate our algorithm on experimentally measured holograms of micrometer-scale colloidal particles, and we show that 20-fold increases in speed, relative to fitting full frames, can be attained while introducing errors in the particle positions of 10 nm or less. The method is straightforward to implement and works for any scattering model. It also enables a parallelization strategy wherein random-subset fitting is used to quickly determine initial guesses that are subsequently used to fit full frames in parallel. This approach may prove particularly useful for studying rare events, such as nucleation, that can only be captured with high frame rates over long times. © 2014 Optical Society of America

OCIS codes: (090.1995) Digital holography; (100.0100) Image processing; (100.2000) Digital image processing; (100.3190) Inverse problems; (100.3200) Inverse scattering; (100.3175) Interferometric imaging.
http://dx.doi.org/10.1364/AO.53.00G177

1. Introduction
Digital holographic microscopy (DHM) is a powerful tool for measuring three-dimensional (3D) dynamics of microscale objects suspended in a fluid. A holographic microscope records a two-dimensional (2D) image, or hologram, resulting from interference between a reference light field and the light scattered from a sample. The 2D hologram encodes phase information that can be used to infer the 3D structure of the sample. Because holograms can be captured with a fast camera, DHM enables measurements of 3D fluctuations and motion on short timescales, making it particularly useful for measuring dynamics of swimming organisms [1–4], living cells [5], and colloidal particles [6,7].

The principal challenge in DHM is recovering the 3D information from the 2D hologram. The most common method is numerical reconstruction, the digital descendant of Gabor’s optical reconstruction [8]. Numerical reconstruction techniques use scalar diffraction theory to backpropagate light from a measured hologram, yielding a 3D image of the sample [9,10]. However, the 3D image can contain artifacts if the sizes of the objects are on the order of the wavelength of light [11,12], as is the case in colloidal suspensions.
An inverse-problem approach can yield more precise and accurate measurements. Ovryn and Izen [13] and Lee et al. [14] measured positions and optical properties of spherical colloidal particles by fitting a Lorenz–Mie scattering solution to holograms, taking the particle position, radius, and refractive index as fitting parameters. Fitting holograms using scattering solutions takes advantage of known information about the particle, such as its shape, and avoids artifacts that lead to systematic errors [15]. This inverse approach has been used for high-precision measurements in fluid dynamics [16] and colloid and interface science [17]. Recent work has extended the fitting technique to clusters of spheres [18,19] and nonspherical particles [20].

However, the fitting technique is computationally intensive. For example, Kaz et al. [17] needed about 100 CPU hours to fit a single particle trajectory consisting of 14,000 frames. For a cluster of particles, the computational burden is even larger: sequentially fitting 14,000 frames of a six-particle cluster [19] would take several CPU months.

How can we reduce the time to fit? One strategy is to parallelize. Although holograms can be processed more quickly using the parallel processing inherent to a graphics processing unit (GPU) [16], it is still difficult to process a set of such holograms in parallel. This is because the nonlinear optimization methods used to fit holograms require initial guesses for the particle position, radius, and index of refraction for each frame. The initial guesses can be obtained from the results of the previous frame but only if the frames are fit sequentially. For clusters containing many particles, automated guesses are difficult because multiple scattering and near-field effects make the hologram sensitive to small changes in particle positions.

A second strategy is to reduce the size of the hologram. Fitting a scattering solution to a hologram is a highly overdetermined problem: typically of order 10 parameters are extracted from a hologram containing of order 10,000 intensity measurements (pixel values). The premise of our approach is that a far smaller number of measurements can sufficiently constrain a fit. This premise was explored by Seifi et al., who used staged resampling [21] and pixel masks [22] to reduce the number of pixels in a hologram and thereby reduce the number of computations required to evaluate the scattering solution. Here we pursue an alternative approach that requires no prior knowledge about the spatial-frequency spectrum of a hologram.

We show that randomly sampling a subset of pixels in the hologram reduces the fitting time by an order of magnitude or more, with little effect on accuracy or convergence. The random-subset approach requires only a single sampling step, is simple to implement, works with any scattering model, and can be combined with GPU computing or resampling methods to reduce computing time even further. The method also allows us to analyze a time series of holograms in parallel. To do this, we first fit a small random subset of each hologram in a time series, then use the results as initial guesses to fit the full holograms in parallel.

2. Background

We first outline the method of fitting scattering solutions to in-line digital holograms, following Lee et al. [14]. A digital hologram $H$ captures phase information through interference of a reference field ($E_{\text{ref}}$) with the field scattered from an object ($E_{\text{scat}}$):

$$
H = |E_{\text{ref}} + E_{\text{scat}}|^2
= |E_{\text{ref}}|^2 + E_{\text{ref}}^* E_{\text{scat}} + E_{\text{ref}} E_{\text{scat}}^* + |E_{\text{scat}}|^2.
$$

(1)

Before fitting, we record the background intensity, which we assume to be equal to the reference wave intensity $|E_{\text{ref}}|^2$. We then calculate a normalized hologram $h$:

$$
h = \frac{H}{|E_{\text{ref}}|^2} = 1 + \frac{E_{\text{scat}}}{E_{\text{ref}}} + \frac{E_{\text{scat}}^*}{E_{\text{ref}}} + \frac{|E_{\text{scat}}|^2}{|E_{\text{ref}}|^2}.
$$

(2)

We extract information about a physical system from $h$ through an inverse-problem approach. We fit a scattering model to $h$ using a nonlinear optimization algorithm; that is, we minimize the objective function

$$
g([x],[y],[z],[n],[r],\alpha \in \mathbb{R})
= |h_{\text{model}}([x],[y],[z],[n],[r],\alpha) - h_{\text{rec}}|^2,
$$

(3)

where the bracketed variables indicate vectors of parameters (one per particle), $h_{\text{rec}}$ is the recorded hologram, and $\alpha$ is an additional fitting parameter needed for convergence [14]. Depending on the shape and structure of the particles, other parameters may be needed. For example, homogeneous nonspherical particles are characterized by their orientations as well as positions. In what follows, we ignore these additional parameters with no loss of generality.

We compute $h_{\text{model}}$ from a light scattering solution. For a single particle we use the Lorenz–Mie solution [14]. For clusters of multiple particles, we use an adaptation [18] of Mackowski’s multisphere superposition solution [23], which takes into account all near- and far-field coupling between the scattered fields. If the numerical minimization of the objective function converges, the result of this fitting procedure is a measurement of the parameters $[x],[y],[z],[n],\text{ and }[r]$.

For larger clusters, we generally use fixed values for the index and radius of the particles, as determined from manufacturer specifications or other experiments. The objective function $g$ therefore simplifies to

$$
g([x],[y],[z],\alpha \in \mathbb{R}) = |h_{\text{model}}([x],[y],[z],\alpha) - h_{\text{rec}}|^2.
$$

(4)

We use local optimization based on the Levenberg–Marquardt algorithm [24] to minimize $g$. Owing to
the large parameter space and narrow minimum, local optimization algorithms fare better than global ones. However, local approaches require an initial guess for the set of parameters and converge only if this guess is close to the minimum of the objective function. Obtaining a sufficiently good guess is difficult. For clusters of spheres, we must manually generate an initial guess for the particle positions, using numerical reconstruction as a guide. However, when studying a time series, we can use the results from fitting one frame as the initial guess for the next, provided that the capture rate is fast relative to particle motion. This approach minimizes manual intervention but precludes processing the frames in parallel.

Fitting is computationally expensive because it requires repeated evaluation of $h_{\text{model}}$, which, like $h_{\text{rec}}$, is an array of intensity values. The objective function $g$ in Eq. (4) is therefore a sum over a large number of pixels:

$$g(x, y, z, \alpha) = \sum_{i,j} (h_{\text{model}}(i, j; x, y, z, \alpha) - h_{\text{rec}}(i, j))^2.$$ (5)

To evaluate $g$, we must calculate $h_{\text{model}}$ at each pixel $(i, j)$ on the detector. This corresponds to calculating the scattered field at a set of locations in space. Calculating the scattered field at each pixel is computationally intensive; in both the Lorenz–Mie and multisphere superposition solutions, the field at each point is given by a series expansion in special functions that themselves require many operations to calculate. On a single core of a 2.67 GHz Intel Xeon X5650 CPU, it takes approximately one second to compute a 200 × 200 pixel hologram of a single 1-μm-diameter particle lying 15 μm from the focal plane. Fitting a single hologram of such a particle takes 15–30 s. It takes much longer to fit the multisphere superposition solution to a hologram of a cluster: about 3 s to compute a hologram of a six-particle cluster of 1.3-μm spheres and over 500 s to fit a single hologram.

3. Approach

Our approach reduces the fitting time by ignoring most of the pixels of a hologram, thereby reducing the number of evaluations required to calculate the objective function $g$. Consider the hologram shown in Fig. 1(a). It contains a large number of pixels, but
only a fraction of these are needed to determine the particle position and properties. For example, much of the information is at low spatial frequencies, as shown in Fig. 1(b). Downsampling techniques like those of Seifi et al. [21] take advantage of the fact that, in many holograms, the largest spatial frequencies are the ones most affected by noise and can be discarded with little effect on the fitted parameters. But even after downsampling [Fig. 1(c)], a hologram retains symmetries and large patches of nearly uniform value, suggesting that it can be compressed further.

Instead of downsampling, we simply choose a random subset of the pixels in the hologram [Fig. 1(d)] and fit a scattering solution to these. This method requires little information about the sample or the optical train, yet is surprisingly effective. The inspiration for this technique comes from compressed sensing, where it has been shown that random sampling is often nearly as efficient as optimal sampling [25–27].

We examined the effectiveness of this technique on two experimentally recorded holograms of polystyrene spheres suspended in water [Figs. 2(a) and 2(d)]. Both holograms we examined are 200 × 200 pixel regions cropped from larger holograms. The first hologram [Fig. 2(a)] was of a 1-μm-diameter sphere and was originally recorded for an in situ viscometry experiment. We modeled this hologram using the Lorenz–Mie scattering solution [14], and we fit the solution to the data using five parameters: three spatial coordinates, the radius, and the index of refraction. The second hologram [Fig. 2(d)] is of a cluster of six 1.3-μm-diameter particles that are nearly touching one another. It was originally recorded for the experiments described by Perry et al. [19]. We modeled this hologram using the multisphere superposition solution [23], and we fit the solution to the data using 18 parameters: three center coordinates for each particle. We fixed the index of each particle to 1.58 and the radius to 0.645 μm. To give the fitting algorithm more freedom to rearrange particles, we allowed the model to adopt configurations where the particles overlapped by up to 10 percent of the radius.

We measured the convergence properties and accuracy of our method as a function of the pixel fraction \( f \) and initial guess quality. The accuracy we report for the fitted parameters is relative to a set of standard values, which we calculated by fitting full frames and averaging the fit results over different initial guesses. The pixel fractions varied from \( f = 1 \), corresponding to the full hologram, to \( f = 1 \times 10^{-3} \). For the single-particle hologram, we chose 10 random subsets at each fraction, and we used an initial guess for \( z \) that differed by 0.2 μm from the standard \( z \), corresponding to a displacement along the optical axis. For the six-particle hologram, we chose 600 random subsets at each fraction, and we used a range of initial guesses created by adding a Gaussian random variable to each standard center coordinate. To vary the guess quality, we increased the width of the Gaussian. We generated 50 initial guesses for each of 12 distributions whose standard deviations ranged from \( \sigma = 1 \times 10^{-4} \) μm to \( \sigma = 1 \) μm. In all cases, we rejected guesses in which particles overlapped by more than 10 percent of their radius.

We used an implementation of the Levenberg–Marquardt algorithm from nmpfit, a Python adaptation of 

\[
\text{LMFIT from MINPACK-1 [28]. We considered a fit to have converged if the sum of squares decreased by less than } 10^{-10} \text{ between iterations, the normalized solution vector changed by less than } 10^{-10} \text{ between iterations, or the cosine of the angle between the solution and all columns of the Jacobian was smaller than } 10^{-10}. \text{ If none of these conditions were met within 100 Levenberg–Marquardt iterations, we halted the minimization and considered the fit to have failed to converge.}
\]

All computations were performed on Harvard’s Odyssey computing cluster. Each fit ran on a single core of a 2.67 GHz Intel Xeon X5650 processor. We used parallel processing only to run independent trials concurrently. Our implementation of random-subset fitting is available in our hologram processing code HoloPy, available as free and open-source software: http://manoharan.seas.harvard.edu/holopy.

4. Results

We find that random-subset fitting extracts particle parameters from holograms more than an order of magnitude faster than fitting the full frame. At pixel fractions near \( f = 1 \), the speedups we obtain increase linearly with decreasing pixel fraction. For both holograms, the speedups plateau at fractions smaller

![Fig. 2. (a), (d) Holograms we use in our study (200 × 200 pixels, 0.12 μm/pixel). (b), (e) Numerical reconstructions showing a plane cutting through the center of the scatterers. (c), (f) Renderings of the structures, based on fitting a scattering model to the holograms. Incident light propagation is into the page in all cases. The inset ball and stick model is rotated to show structure.](image-url)
than \( f = 0.01 \) (Fig. 3): the plateau is at 100× speedup for the single-particle hologram and 25× for the hologram of the six-particle cluster. The plateau occurs because the time required to compute scattering coefficients is independent of the pixel fraction. This overhead is larger for the multisphere superposition solution, explaining the lower plateau for the six-particle cluster.

We also find that fitting such small fractions does not require better initial guesses. For the full hologram of the six-particle cluster, the fitting algorithm converges for initial guesses within about 100 nm (per coordinate) from the standard values, as shown in Fig. 4. Fits converge with similar reliability for pixel fractions as low as \( f = 0.01 \). Reducing the fraction further reduces the probability of convergence, but there is little reason to use such small subsets given the diminishing returns on speedup.

Finally, we examine accuracy. In a simulated case with a known correct answer (Fig. 5 dotted line), we find that we can recover the correct particle positions to within numerical roundoff error, even at subset fractions as low as \( f = 1 \times 10^{-4} \) (which correspond to only 20 pixels). Thus, in the absence of noise, the fit accuracy is limited only by numerical precision and minimizer cutoff. With our experimentally measured holograms, we find a larger deviation between the fitted parameters and the standard values at fractions smaller than the full frame \( f = 1 \) (Fig. 5). However, the errors are comparable to the intrinsic measurement error, which is approximately \( 10^{-3} \) \( \mu m \) for single particles [17] and \( 10^{-2} \) \( \mu m \) for clusters [29]. Thus, we expect that fitting to subsets as small as \( f = 0.1 \) will not impact accuracy.

5. Discussion

Our results show that fitting scattering models to random subsets of holograms offers large speedups with few tradeoffs, either in convergence or accuracy. Two questions arise: How should it be used as part of a workflow for analyzing holograms? And why is random sampling so effective?

Given how simple it is to apply random sampling, there seems little reason not to use it when fitting holograms. Some experimentation may be required to find the appropriate subset fraction, which varies with the scatterer and experimental conditions. However, we find that using 10 percent of the pixels generally gives good results.

We have found the method to be particularly useful for analyzing time series of holograms quickly and accurately. We use a two-stage procedure, the first stage of which is a serial “rough fit” of each hologram in the time series, using a fraction of 0.01 or 0.1. We also limit the minimization algorithm to two iterations. The rough fit completes in approximately 1/200th the time required to serially fit the full frames: We gain a factor of 20 by fitting to small subsets and
another factor of 10 by limiting the number of iterations. In the second stage, we use the results from the first stage as initial guesses to fit the full holograms. Because these fits can be done in parallel, the speedup scales linearly with the number of nodes.

As to the second question, on the effectiveness of the technique, we can offer only an intuitive justification. Consider a linear transformation, operating on a hologram, that compresses all the information about particle parameters into a small number of basis elements. We expect this transformation to depend on the shape, size, and optical properties of the scatterer. Hence we do not—and quite possibly cannot—know either the transformation or the optimal basis. However, for certain classes of signals, it has been shown that random sampling allows efficient compression without knowledge of the optimal basis [25–27]. We cannot prove that a hologram is in this class of signal, but the success of our method suggests that it is close enough for practical purposes, at least for holograms of spherical particles or clusters of spheres. Future work might yield a more rigorous mathematical justification of this idea.

6. Conclusions

We have shown that the random-subset technique speeds up fitting by a factor of 20 or more, with little effect on accuracy or convergence compared to fitting to the full hologram. The technique is simple to implement and works for different scattering models. It can be used in concert with GPU or CPU-based parallel processing, with other algorithmic improvements, or with faster implementations of scattering models. In our calculations, we sampled from a uniform distribution of pixel locations, but nonuniform sampling may allow even greater speedups by weighting areas of a hologram most useful for constraining a fit. In future work, it will be particularly interesting to determine if there are nonuniform distributions that are effective for a wide variety of scatterers.

Because recovering precise 3D information from measured data is the most time-consuming step in holographic microscopy, the method we have demonstrated should allow faster turnaround in experiments and may enable new experiments that were previously infeasible. The ability to process large quantities of data more quickly will be particularly useful for studying rare events, such as nucleation, that can only be captured with high frame rates over long times.

This work was funded by the National Science Foundation (NSF) through Grant No. DMR-1306410 and by the Harvard MRSEC through NSF Grant No. DMR-0820484. Scattering calculations and hologram fitting were performed on the Odyssey cluster, managed by the FAS Science Division Research Computing Group at Harvard University. We thank Ameya Agaskar for helpful discussion on the mathematical background of this technique. Our nmpfit routine comes from STSCI_PYTHON, a product of the Space Telescope Science Institute, which is operated by AURA for NASA.

References


