Three-dimensional reconstruction of particle holograms: a fast and accurate multiscale approach.
Mozhdeh Seifi, Corinne Fournier, Loïc Denis, Delphine Chareyron, Jean-Louis Marié

To cite this version:

HAL Id: ujm-00744235
https://hal-ujm.archives-ouvertes.fr/ujm-00744235
Submitted on 23 Oct 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
3D reconstruction of particle holograms: a fast and accurate multi-scale approach

Mozhdeh Seifi,¹ Corinne Fournier,¹ Loic Denis,¹ Delphine Chareyron²,³ and Jean-Louis Marié²

¹ Université de Lyon, F-42023, Saint-Etienne, France, CNRS, UMR5516, Laboratoire Hubert Curien, Université de Saint-Etienne
   Telecom Saint Etienne, F-42000, Saint-Étienne, France,
² Laboratoire de Physique de l’ENSEL, UMR5672 CNRS, 46 Allée d’Italie, 69364 Lyon Cedex 07, France,
³ Laboratoire de Mécanique des Fluides et d’Acoustique UMR5509, Ecole Centrale de Lyon - CNRS - Université Claude Bernard Lyon 1 - INSA Lyon, 36 avenue Guy de Collongue, 69134 Ecully cedex, France.

corinne.fournier@univ-st-etienne.fr

In-line digital holography is an imaging technique which is being increasingly used for studying 3D flows. It has been previously shown that very accurate reconstructions of objects could be achieved with the use of an inverse problems framework. Such approaches however suffer from higher computational times compared to less accurate conventional reconstructions based on hologram back-propagation. To overcome this computational issue, we propose a coarse-to-fine multi-scale approach to strongly reduce the algorithm complexity. We illustrate that an accuracy comparable to state-of-the-art methods can be reached while accelerating parameter-space scanning. © 2012 Optical Society of America

OCIS codes: 100.3190, 090.1760, 100.5010, 100.6640, 100.2000

1. Introduction

Study of 3D flows has many applications in different fields of science such as fluid mechanics (e.g., study of turbulence [1], droplet evaporation [2,3]) and biology (e.g., study of locomotion of micro-organisms [4]). The goal of these applications is mainly to study motion of turbulent 3D fluid structures or the dispersion and mixing of particles inside a volume. The impact
of the micro-particles suspension on the rheology is interesting in the context of non-Newtonian fluids. Other examples are study of the interactions between micro-particles. For these applications it is therefore essential to use imaging tools to track 3D positions of these micro-particles and/or to observe the changes in their shapes using a sequence of measurements. The requirements of such methods are mainly high accuracy, fast imaging techniques and simplicity of the image processing tools for users.

Since its early developments, in-line digital holography (DH) has increasingly been used in studies of 3D flows (see [7–10] and references therein) because of its following advantages: high speed imaging technique, very simple setup, accurate measurements (without magnification the accuracies are 1µm for transversal and 10µm for depth estimation, with magnification they can reach up to 5nm and 100nm respectively). DH consists of two steps. First, during the recording step, objects are illuminated by a laser beam and diffraction patterns are captured by the camera. This 2D hologram is then processed in a second step to extract 3D locations, and possibly shape information, of the illuminated objects.

Most hologram processing techniques in the literature are based on hologram diffraction. They first simulate optical reconstruction of the object wave (using, for example, the Fresnel transform to back-propagate the hologram). Then, they segment the obtained 3D volume to detect and locate in-focus objects. Let us note that hologram back-propagation does not fully invert hologram formation, which is the reason for several artifacts encountered in reconstructed volumes such as ghost images, border effects and twin images.

Segmentation of in-focus objects is performed based on the real part or imaginary part of reconstructed volume or some transformation of this reconstructed object field (e.g., wavelet transform or integrated amplitude modulus). From a signal processing perspective, hologram-diffraction based techniques have limited accuracies due to (i) signal truncation which dramatically limits the accuracy of field reconstruction close to the borders, and (ii) low spatial resolution of digital sensors which either results in false object detection (due to ghost images) or enforces limitations on the recording setup leading to a lower signal magnitude, and consequently lower signal-to-noise ratio (see [18] for details).

In contrast to classical hologram-diffraction approaches, methods for reconstructing particle holograms based on an inverse problems formulation lead to optimal detection performance and efficient estimators. These methods are sometimes referred to as compressive sensing methods in the literature. Starting from a model of the diffraction pattern of a particle, these approaches search for the location and size of particles that best “explain” observed data (i.e., maximum likelihood estimation of location and size). In addition to higher accuracy of estimations, such inverse problems approaches can expand the field of view outside of the sensor area up to a factor 16, which is out of reach of classical methods.
An important practical issue of these approaches, though, is their huge computational cost. We suggest using a coarse-to-fine scheme with a multi-scale (not necessarily dyadic) pyramid to overcome this limitation. Connections between multi-scale techniques and holography have long been made [16,23,24]. We show that parameter space scanning can be made time-efficient using multiple scales by first exploring exhaustively a small sub-volume, and then focusing computational effort on local optimization with refined resolutions.

The structure of the paper is as follows: Sec. 2 gives a brief introduction to inverse problems approaches in DH, then Sec. 3 describes the main body of this paper: In Sec. 3.A the proposed multi-scale approach is described in detail. In Sec. 3.B the analytical approximation of the diffraction pattern model of spherical particles is formulated. In Sec. 3.C and 3.D we discuss the choice of pyramid height and a criterion for adaptively stopping local optimizations at each scale. We then compare both speed and accuracy to the existing inverse problems technique in Sec. 4.

2. Inverse problems approaches in digital holography

Inverse problems represent a general class of problems where unknowns are related to measurements through a known model (simulating the measurements is referred to as the “direct problem”), and estimating the unknowns from their corresponding measurements can be a difficult task.

In the context of DH reconstruction, two kinds of problems can be considered: (i) non-parametric reconstruction of objects; (ii) detection and location of simple parametric shapes (typically, spherical particles).

Non-parametric reconstruction of objects requires inversion of the diffraction operator which maps a 3D transmittance volume to a 2D hologram. This is an ill-posed problem that requires regularization for its inversion. Successful reconstructions have been obtained with smoothness constraints (enforced by total variation minimization [25,26,28–30]) or sparsity in the spatial domain [22].

Parametric reconstruction is useful for particle hologram reconstruction. We will focus in the following on spherical particles which is important in many metrological applications. The problem in that case is to estimate 3D location and diameter of particles from a single hologram. It has been shown [31] that a particle hologram \(d\) can be well approximated as the sum of the diffraction pattern \(m_i\) of each particle:

\[
d \approx \sum_{i=1}^{N_{\text{part}}} \alpha_i m_i,
\]

where hologram \(d\) and models \(m_i\) are adequately centered (i.e., constant offset removed). We represent both the \(N\)-pixels hologram and diffraction patterns as vectors in \(\mathbb{R}^N\), as is
common practice in the context of linear inverse problems. \(N_{\text{part}}\) represents the number of particles in the hologram. The model \(m_i\) of the diffraction pattern created by the \(i^{\text{th}}\) particle is a non-linear parametric model (see Eq. 7). The coefficients \(\alpha_i\) are positive values that account for the amplitude of each diffraction pattern (this amplitude may vary from one particle to another due to inhomogeneity of the incoming wave).

In Eq. 1, while the diffraction patterns created by each particle are modeled through the models \(m_i\), the interferences between the particles’ diffraction patterns are approximated by an incoherent summation. This linearization of the hologram formation model is essential in in-line holography and is valid when particles are small (i.e., with diameters less than a few hundred microns) and not too numerous. Let us note that an empirical criterion proposed by Royer [27] states that the holograms are of good quality provided that the projected surface of the particles are less than 1\% of the sensor surface. In the rest of this paper, we will use the linear model under the assumption that the inter-particle interferences are negligible. This assumption is widely and successfully used in the literature of in-line digital holography of micro-particles [19, 31, 39, 40].

Accurate estimates can be obtained by weighted least-squares fitting [19, 21], i.e., by finding parameters of a single diffraction pattern model \(m\) that minimize the square Mahalanobis distance to the data considered as the cost function:

\[
D^2_W(d, m) = (d - m)^t W (d - m)
\]

with \(W\) the inverse of the covariance matrix of noise. Even if noise is considered to be uncorrelated, the matrix \(W\) plays an important role since it represents the hologram support. \(W\) is then diagonal, with diagonal entries equal to 0 for unmeasured pixels (e.g., pixels outside of the hologram support) and equal to the inverse of noise variance \(1/\sigma^2\) for measured pixels. When the noise is Gaussian, then maximum likelihood estimates of the parameters of model \(m\) are obtained by minimizing \(D^2_W(d, m)\). To ensure robust detection with holograms of several particles, an iterative detection/localization scheme has been proposed in [31].

To have more concise mathematical expressions, we introduce the scalar product of two vectors \(\langle u, v \rangle_W\) and the induced norm \(\|u\|_W^2\) as

\[
\langle u, v \rangle_W = \frac{u^t W v}{1^t W 1}, \quad (\text{for a diagonal } W: W = \text{diag}(w))
\]

\[
\|u\|_W^2 = \langle u, u \rangle_W = \frac{u^t W u}{1^t W 1}, \quad (\text{for a diagonal } W: W = \text{diag}(w)).
\]

Using this notation, minimizing Mahalanobis distance is equivalent to minimizing \(\|d - m\|_W^2\). The 3D reconstruction algorithm can then be summarized in three steps [19, 31]:
1. An exhaustive search in parameter space to find the best matching element in a discrete dictionary of diffraction patterns \{m_1, \ldots, m_K\} (Fig. 1(b)). The goal of this step is to identify the size \(r\) and 3D location \((x, y, z)\) of a single particle that would best explain the data (i.e., particle with maximum likelihood).

\[
\arg \min_{\alpha \geq 0, 1 \leq i \leq K} \| \alpha m_i - d \|_W^2
\]

Partial minimization with respect to \(\alpha\) leads to a generalized maximum correlation criterion [12]:

\[
\arg \max_i \frac{[\langle d, m_i \rangle_W]^2}{\| m_i \|_W^2}
\]

where \([\cdot]_+ = \max(\cdot, 0)\) denotes the positive part and is used to eliminate anti-correlated models from consideration.

2. A local optimization step which improves the estimation of the particle location and size obtained at the previous step. This is done by numerical optimization of the parameters of non-linear model \(m(x, y, z, r)\). This step overcomes the accuracy limitation induced by the sampling used to generate the discrete dictionary \{m_1, \ldots, m_K\}. Sub-pixel accuracy is reached by continuous optimization of the fitting criterion (Fig. 1(c)): The initial values of parameters \(x, y, z, r\) used to begin local optimization are the values \(x_i, y_i, z_i, r_i\) corresponding to sampled model \(m_i\).

3. A “cleaning” step where observed data is updated by removing the diffraction pattern of the particle detected and located previously. Processing steps 1 to 3 are then repeated on the residuals (i.e., data with previously detected objects removed).

This algorithm is described in more detail by Soulez et al. [19, 31] and is based on the CLEAN algorithm introduced by Högbom in radio-astronomy [32] and matching pursuit by Mallat and Zhang [33]. The use of a local optimization step can be seen as a way to handle a continuous dictionary of diffraction patterns. This algorithm belongs to the family of greedy algorithms [34].

It should be noted that one advantage of having a parametric image formation model is the possibility of using estimation theory to calculate lower bounds on the achievable accuracies of parameter estimation (see Sec. 3.D). These lower bounds can help to assess performance of different hologram processing algorithms. According to estimation theory [35], the accuracy on parameter estimation by maximum likelihood asymptotically reaches these lower bounds (with large numbers of measurements). Based on maximum likelihood estimation, the described inverse problem approach can thus be considered optimal in this respect.
Fig. 1. (a) Hologram of a spherical micro-particle, (b) sampled 4D search space for exhaustive search step, (c) a 1D profile of the cost function of particle for visualization purpose. This cost function is jointly optimized in practice over the four parameters \((x, y, z, r)\) during the local optimization step. The result of the exhaustive search is used as the initial point for this optimization. The parameters used in this example are the following: laser wavelength \(\lambda = 0.532\ \mu m\), pixel size 7 \(\mu m\), signal-to-noise ratio 50, size of the hologram \(2048 \times 2048\) pixels, distance between particle and sensor 0.3 m and radius 33 \(\mu m\).
However, high accuracy comes at the cost of large computation times which can be considered a strong disadvantage compared to classical approaches. In order to reduce this problem, we propose a multi-scale approach to dramatically accelerate step 1 of greedy particle detection algorithm. This approach is introduced in detail in the next section.

3. Proposed multi-scale approach to 3D particle localization and sizing

The inverse problems approaches presented in previous section model both hologram formation (particle diffraction, hologram sampling and finite support) and objects (opaque spherical particles defined by their 3D coordinates and radius). They lead to optimal detection and estimation performance at a large computational cost. The exhaustive-search step (step 1 described in Sec. 2) can be made much faster by down-sampling the hologram. This leads to a rough estimate of particle size and position that can further be improved by local optimization (i.e., model fitting) on increasingly higher-resolution versions of the hologram.

We first give an overview of the method, then detail how the hologram multi-scale pyramid is built, and design adaptive stopping criteria.

3.A. Overview of the method

The exhaustive-search step requires exploring a sampled 4D parameter space. To reach pixel-accuracy in \((x, y)\) and sufficient accuracy in \((z, r)\), hundreds of \((z, r)\) pairs may need to be considered for each \((x, y)\) location, leading to hundreds of millions or billions of quadruples \((x, y, z, r)\) to be tested. Shift-invariance of the model can be exploited by using the Fast Fourier Transforms (FFT). The search is thus reduced for each exhaustive-search step to the computation of hundreds of convolutions to evaluate the generalized maximum correlation criterion given in Eq. 6 (the criterion requires 7 FFTs for each \((z, r)\) pair [19]). This is then repeated for each particle unless multiple particle detection is implemented [12,36]. To further reduce computational complexity, we propose to carry out the exhaustive search on a down-sampled version of the hologram, as described in algorithm FAST (Fig. 2).

Before getting into the detail of each step, here is a sketch of the algorithm. Since exhaustive search is a computational bottleneck, we build a multi-resolution pyramid from the hologram (see Fig. 3(a)) and perform an exhaustive search on the coarsest scale only. Local optimization is then performed on increasingly fine scales, restarting numerical optimization each time from the parameters obtained at the previous (coarser) scale. The down-sampled hologram at level \(k\) is computed by low-pass filtering and down-sampling the full-resolution hologram \(d\) by a linear filter \(F^{(k)}\) (see Sec. 3.C.1) where \(T_k\) corresponds to the period of down-sampling. We denote down-sampled holograms as \(\tilde{d}\) in algorithm FAST (Fig. 2).

Using a coarse resolution hologram for the exhaustive search step not only reduces the number of \((x, y)\) samples by a factor \(\frac{T_k^2}{2}\) (and, thus, the size of images on which 2D FFTs are
computed) but also makes the cost function $D_W^2$ smoother. Sampling of parameters $z$ and $r$ (i.e., depth and radius of a particle) can also be made coarser in this way. Fig. 3(a) illustrates the widening of cost function $D_W^2$ when coarser resolution holograms are considered (a profile of $D_W^2$ along axis $z$ is drawn). The risk of getting trapped in a local minimum is then much weaker, which relaxes sampling constraints that guarantee being within reach of the global minimum. We discuss convergence of our multi-resolution algorithm to the global minimum in Sec. 3.E.

We now detail how each step of the algorithm is performed in order to reach an accuracy comparable to that of the slower single-resolution approach. We begin with a description of the diffraction-pattern model $m$ in Sec. 3.B, then describe down-sampling filtering and the choice of the maximum down-sampling period in Sec. 3.C. We detail the stopping criteria for each refinement step in Sec. 3.D.

3.B. Diffraction pattern model

A spherical opaque particle generates a diffraction pattern made of concentric rings with both frequency and amplitude modulations. For particles small with respect to the recording distance ($\pi r^2 \ll \lambda z$), the diffraction pattern is well modeled by a linear frequency modulation (chirp) and a cardinal Bessel amplitude modulation [21, 31, 37]. The use of a digital camera generates an integration effect over a pixel area which introduces additional amplitude modulation by cardinal sines. The model $m_i$ of a particle located at coordinates $(x_i, y_i, z_i)$ with radius $r_i$ can then be written [18]:

$$m_i(\ell) = \frac{\pi r_i^2}{\lambda z_i} \cdot \sin \left( \frac{\pi \rho_{\ell}^2}{\lambda z_i} \right) \cdot J_{1c} \left( \frac{2\pi r_i \rho_{\ell}}{\lambda z_i} \right) \cdot \text{sinc} \left( \frac{\pi s \Delta x_{\ell}}{\lambda z_i} \right) \cdot \text{sinc} \left( \frac{\pi s \Delta y_{\ell}}{\lambda z_i} \right)$$

where $m_i(\ell)$, the $\ell^{th}$ element of vector $m_i$, represents the $\ell^{th}$ pixel of the diffraction pattern of particle $i$. Particle $i$ is located respectively at distances $\Delta x_{\ell}$ and $\Delta y_{\ell}$ from the $\ell^{th}$ pixel along $x$-axis (resp. $y$-axis). The radial distance $\rho_{\ell} = \sqrt{\Delta x_{\ell}^2 + \Delta y_{\ell}^2}$ corresponds to the distance between the center of the diffraction rings (i.e., the projection of the center of particle $i$ onto plane $(x,y)$) and the center of pixel $\ell$. The wavelength of the laser is written $\lambda$, and $s^2$ is equal to the sensitive area of a pixel (i.e., fill-factor times the area of a pixel). The cardinal Bessel function of the first kind is written $J_{1c}(\cdot)$ while sinc$(\cdot)$ denotes the cardinal sine function.

3.C. Construction of the multi-resolution pyramid

3.C.1. Filtering and down-sampling

The hologram at level $k$ is obtained by application of linear filter $F^{(k)}$ on the original hologram $\hat{d}$:

$$\hat{d} = F^{(k)}d$$

(8)
Algorithm FAST (Fast and Accurate multi-Scale esTimator of size and location of particles)

Input: hologram \( d \), parameters of the setup (laser wavelength, pixel size, range of plausible particle locations and radii \( T \)).

Output: estimated parameters of the \( N_d \) detected particles: \( \{x_i, y_i, z_i, r_i\}_{i=1,...,N_d} \)

choose a down-sampling factor \( k_{\max} \)

repeat
  \( \tilde{d} = F^{(k_{\max})} d \) \( \triangleright \) see Sec. \( 3.C.2 \)
  \( i^* \leftarrow \arg \max_i \left[ \left( \frac{\langle \tilde{d}, \tilde{m}_i \rangle}{\| \tilde{m}_i \|^2_W} \right)^2 \right] \)
  \( (x, y, z, r) \leftarrow (x_{i^*}, y_{i^*}, z_{i^*}, r_{i^*}) \)
  \( k \leftarrow k_{\max} \)
  while \( k \geq 1 \)
    \( \tilde{d} = F^{(k)} d \) \( \triangleright \) particle detection loop
    \( (x, y, z, r) \leftarrow \arg \min_{x,y,z,r,\alpha} D_W^2(\tilde{d}, \tilde{m}) \) \( \triangleright \) down-sample hologram \( d \) by the period \( T_{k_{\max}} \)
    \( (x, y, z, r) \leftarrow \arg \min_{x,y,z,r,\alpha} D_W^2(\tilde{d}, \tilde{m}) \) \( \triangleright \) do exhaustive search on coarsest scale
    \( x, y, z, r \leftarrow x_i^*, y_i^*, z_i^*, r_i^* \) \( \triangleright \) set obtained parameters as initial values
    \( k \leftarrow k - 1 \) \( \triangleright \) refine particle parameters, stop according to Sec. \( 3.D \)
    \( T_k \leftarrow \left\lfloor T_{k+1}/2 \right\rfloor \) \( \triangleright \) descend on pyramid
  end while
  if \( \alpha > \alpha_{\text{min}} \) and \( (x, y, z, r) \in T \) \( \triangleright \) parameters fulfill conditions for being a particle
    store \( (x, y, z, r) \) \( \triangleright \) add parameters to the collection of detected particles
  else
    stop \( \triangleright \) return already detected particles and finish
  end if
end repeat

Fig. 2. The proposed multi-scale algorithm for particle detection and sizing from a digital hologram.
Fig. 3. (a) Schema of the proposed multi-scale algorithm, (b) 1D profile of the cost function computed on the original hologram: black crosses show the results of estimation after each step of pyramidal multi-scale algorithm on the profile of the cost function, black circle shows an example of coarse estimation from exhaustive search of FAST (Fig. 2) with $k_{max} = 0$ (single-scale approach). As shown here, this coarse detection should be found inside the main basin of the cost function whereas the coarse estimation using pyramidal multi-scale algorithm could be outside the basin.
where $F^{(k)}$ is a $(N/T_k^2) \times N$ matrix transforming the $N$-pixels hologram $d$ into a coarser scale hologram $\tilde{d}$ with $T_k^2$ times less pixels.

To prevent severe aliasing effects, $F^{(k)}$ must have low-pass behavior. We chose to average $T_k^2$ values together before down-sampling to reduce aliasing while leaving noise uncorrelated (so that matrix $W$ in Eq. 2 and following remains diagonal). In the following, $F^{(k)}$ is an averaging filter followed by a down-sampling operation. Other choices for $F^{(k)}$ are suggested in Sec. 5.

3.C.2. Maximum down-sampling period selection

The maximum down-sampling period $T_{k_{\text{max}}}$ defines the resolution for which exhaustive search is performed before successive refinements by local optimization are done (see algorithm FAST in Fig. 2). Two reasons for not selecting an arbitrarily large down-sampling factor are: (i) using lower resolution holograms reduces the accuracy, especially on $z$ and $r$, and may lead to initial estimates of parameters $(x, y, z, r)$ too far from the actual 3D location and size of the particle to find the correct parameters by successive refinements; (ii) down-sampled versions of the model are known in closed form provided the down-sampling factor is limited.

We address the first issue by requiring that the down-sampling period $T_{k_{\text{max}}}^{(i)}$ is such that a significant number $q$ of diffraction fringes are still visible on coarse-scale hologram $\tilde{d}$. We derive in appendix A a bound for $T_{k_{\text{max}}}^{(i)}$ and $k_{\text{max}}$ as:

$$T_{k_{\text{max}}}^{(i)} = \left\lfloor \frac{1 \kappa \sqrt{\lambda z_{\text{min}}}}{2q + 1/2} \right\rfloor, k_{\text{max}}^{(i)} = \left\lfloor \log_2(T_{k_{\text{max}}}^{(i)}) \right\rfloor$$

with $\lambda$ the wavelength, $z_{\text{min}}$ the a priori minimum depth of a particle, $\kappa$ the pixel width (pixels are assumed to be square for notational convenience) and the brackets representing the floor function. We discuss in Sec. 3.E and show on our experiments in Sec. 4 that setting $q$ equal to 10 is enough to obtain correct estimates of particle parameters.

The second issue relates to the ability to express the down-sampled models in closed form. This is essential for fast estimation of Mahalanobis distance between coarse holograms and down-sampled models during local optimization. We recall from appendix B that hologram convolution with a kernel of limited size amounts to an amplitude modulation of the diffraction rings formulated by Fresnel function. This is due to the correspondence between frequencies of diffraction rings and (spatial) radii of these fringes. The coarse-resolution model $\tilde{m}_i$ is then obtained by weighting the original model expression:

$$\tilde{m}_i(\ell) = m_i(\ell') \cdot \tilde{f}(\ell')$$

where $\tilde{m}_i(\ell)$ is the $\ell^{th}$ pixel of coarse-resolution model $\tilde{m}_i$, $\ell'$ is the index of the corresponding pixel in full-resolution model, and $\tilde{f}$ is the Fourier transform of the low-pass filter (an
averaging filter in our case) used to build the pyramid (see Appendix B). Computation of the coarse-resolution model $\hat{m}_i$ at level $k$ (which contains $N/T_k^2$ pixels) requires only $N/T_k^2$ evaluations of model $m_i$ and frequency responses $\tilde{f}$ (i.e., far less than $N$ evaluations for the full-resolution hologram/model).

To derive Eq. (10) it is assumed that the filter and the aperture of the objects are real, symmetric with respect to origin and small. This last assumption defines a second constraint $T_{k_{max}}^{(ii)}$ on the down-sampling period (see Appendix B: Ineq. (16)). Considering a maximum a priori radius for particles $r_{max}$, and sensor size $L$ (hologram width), $T_{k_{max}}^{(ii)}$ and $k_{max}$ are (see Appendix C for details):

$$T_{k_{max}}^{(ii)} = \left\lfloor 2 \left( \sqrt{\frac{L}{10}} - \frac{r_{max}}{\kappa} \right) \right\rfloor, \quad k_{max}^{(ii)} = \left\lfloor \log_2(T_{k_{max}}^{(ii)}) \right\rfloor$$

In order to satisfy both conditions we choose the down-sampling factor as: $\min(T_{k_{max}}^{(i)}, T_{k_{max}}^{(ii)})$. Fig. 4 shows the effect of down-sampling using only Eq. (11) as well as considering both criteria in Eq. (9) and Eq. (11).

3.D. Stopping criteria for successive local optimizations

A previous study [38] has shown that resolution in digital holography could be estimated by computation of Cramer-Rao lower bounds (CRLBs). This approach can be extended to the
evaluation of the accuracy of 3D particle location and size estimates. Let vector \( \theta \) denote the vector of particle parameters: \( \theta = (x, y, z, r)^t \). Cramer-Rao lower bounds express the variance of the \( a^{th} \) parameter estimate \( \hat{\theta}_a \) based on the inverse of Fisher information matrix \( I \):

\[
\text{var}(\hat{\theta}_a) \geq \left[I^{-1}\right]_{a,a}
\]

with Fisher information matrix defined by [38]:

\[
[I]_{a,b} = \frac{1}{2} \left( \frac{\partial m(\theta)}{\partial \theta_a} \right)^t W \left( \frac{\partial m(\theta)}{\partial \theta_b} \right)
\]

If noise is white and stationary, \( W \) is equal to \( 1/\sigma^2 \) times the identity matrix. Fisher information matrix can be easily derived for resolution level \( k \) of the multi-resolution pyramid:

\[
[I]^{(k)}_{a,b} = \frac{T^2_k}{2\sigma^2} \left( \frac{\partial \tilde{m}(\theta)}{\partial \theta_a} \right)^t \left( \frac{\partial \tilde{m}(\theta)}{\partial \theta_b} \right)
\]

Using Eq. 12 and Eq. 14, the standard deviation on parameter estimation can be simply calculated and used as the stopping criteria for consecutive local optimization steps.

3.E. Convergence

Maximum likelihood estimation of particle parameters requires minimization of the cost function (i.e., Mahalanobis distance) \( D^2_W(\mathbf{d}, \mathbf{m}) \). Finding the global minimum of this cost function is difficult due to non-convexity of the criterion. Local optimization starting from an initial guess \( \theta^0 \) will generally lead to the global optimum \( \theta^* \) only if \( \theta^0 \) is already in the convexity region (i.e., the main basin of the cost function) \( B(\theta^*) \). The exhaustive search step needs to perform a dense enough sampling of parameters space to ensure that region \( B(\theta^*) \) is probed.

As noted in Sec. 3.A, low-pass filtering applied to the hologram to produce the coarse-resolution levels of the multi-resolution pyramid gives better behaved cost-functions (i.e., with smoother and larger main basin). This effect is visible on profiles along the \( z \) axis of cost functions plotted in Fig. 3(a). Successive refinements obtained by local optimization of the cost function on progressively finer resolutions also helps to find the global minimum \( \theta^* \) even if the initial guess obtained on the coarsest scale \( \theta^0 \) is not inside the basin \( B(\theta^*) \): Fig. 3(b). The global convergence condition \( \theta^0 \in B(\theta^*) \) is relaxed into a sequence of convergence conditions at each scale:

\[
\begin{cases}
\theta^0 \in B(\theta^*_{(k_{\text{max}})}) \\
\theta^*_{(k_{\text{max}})} \in B(\theta^*_{(k_{\text{max}}-1)}) \\
\vdots \\
\theta^*_{(1)} \in B(\theta^*)
\end{cases}
\]

13
We show in the next section that, with our choice of $k_{\text{max}}$, these conditions are fulfilled in practice.

4. Experiments and results

To quantify the performance of the proposed multi-scale algorithm for particle detection and sizing in DH, simulated and real holograms were processed using the standard inverse problems approach and our multi-scale algorithm. We used the Matlab as the programming environment. The reported time gain results are obtained using the FFTW library \cite{12} and OpenMP to exploit multi-threading on a 6-core CPU for the calculations of the forward and backward Fourier transforms and the models required for the optimization steps. In addition, a close estimation of the models are calculated by interpolating the values of a precomputed table to reduce the calculation time.

This section presents the results obtained from simulated and experimental holograms. Simulations are performed in two different test cases: (i) holograms of particles with various radii placed at different depth positions; (ii) holograms of the same particles with different noise levels. Speed and accuracy are compared to the reference single-scale approach. The hologram processing method is then validated on experimental holograms from a recent study \cite{39}.

4.A. Simulations

To study the multi-scale approach on particles with different radii and different depth position, two sets of simulations were performed to estimate (i) the speedup brought by our multi-scale approach; (ii) the accuracy of our algorithm compared to the standard single-scale inverse problems approach.

For the first case, simulations were performed for 100 holograms each containing 5 particles with randomly chosen coordinates. We used the following parameters: particles with radii between 20 $\mu$m and 70 $\mu$m were placed at distances ranging from 30 cm to 48 cm of a 1024 × 1280 pixel camera with pixel size of 21.7 $\mu$m and fill-factor of 0.84. White Gaussian noise was added, leading to a SNR of $\approx$ 16 (SNR is calculated as the ratio of the magnitude of signal over the standard deviation of noise). We choose, to be in the same conditions as the experiments with the magnification of 1.42. The maximum down-sampling factor calculated from Eq. 9 and Eq. 11 for preserving 10 fringes ($q = 10$) was $T_{k_{\text{max}}} = \min \{9, 4\} = 4$ and so $k_{\text{max}} = 2$.

The inverse problems approach was then used to process holograms with and without (i.e., $k_{\text{max}} = 0$) the multi-scale approach. The results show accurate particle detection and estimation for both configurations. The RMSE of the particle coordinate estimates as well as the computational time costs are shown in Table. The accuracies indicated in this table are
calculated for numerical simulations and are therefore better than achievable accuracies of the algorithm on real data. As indicated in this table, the multi-scale algorithm has virtually the same accuracy for $k_{\text{max}} = 2$ as for the single-scale case ($k_{\text{max}} = 0$). However the time cost of the first one is more than four times less than the second one, which justifies the interest in using the pyramidal multi-scale algorithm. According to our results shown in this table, the time costs of local optimization step of both algorithms is the same. The comparison between the exhaustive search step of single-scale approach with time cost of multi-scale algorithm excluding the last local optimization shows the time gain equal to a factor of 10. The overall time gain was on average equal to a factor of 4.

The second set of the simulations was dedicated to the study of the possible limitations of multi-scale approach. Apart from the limitations on the number of the pyramid’s levels (which is explained in Sec. 3.C.2), we performed several simulations on 100 holograms decreasing the SNR to check the limitations of our multi-scale method. In all our experiments, the results indicate that where ever the one-scale approach succeeds in the detection and parameters estimation, so does the multi-scale approach. This result is not surprising since the filtering on the upper levels of multi-scale approach does not change the noise characteristics of the global least squares fitting problem. We found a lower bound of 0.3 on the SNR for the previously used application parameters.

4.B. Experimental holograms

To ensure that the performance of the proposed algorithm remains the same for real data, it was also tested on experimental holograms of mono-dispersed water droplets [39]. One captured hologram is shown in Fig. 5(a). The droplets were generated by a piezoelectric jetting device manufactured by MicroFab Technologies. This injector produces close to mono-dispersed droplets with radii of 31 µm ± 0.5 µm. The other experimental parameters are the same as for the simulations except for the SNR (≈ 6). To process these holograms we use a range of [27.5 - 32.5] µm for radii. The qualitative assessment of results shows low residuals in the cleaned holograms (see Fig. 5) and the quantitative performance assessment was carried on to validate the pyramidal multi-scale algorithm. As for the simulated holograms, the accuracy of estimation for multi-scale algorithm was the same order of magnitude as for the single-scale inverse problems approach i.e., the discrepancy between estimation with single-scale and multi-scale was smaller than the achievable accuracy (10^{-8} m for x and y, 10^{-6} m for z and 10^{-8} m for r)). In this case, the time gain using multi-scale algorithm (Fig. 2) was a factor of 3.
Fig. 5. (a) Experimental hologram of six spherical micro-particles in the field of view of sensor, (b) cleaned hologram with FAST (Fig. 2) for $k_{\text{max}} = 2$, (c) cleaned hologram with FAST (Fig. 2) for $k_{\text{max}} = 0$ (single-scale approach). In the captured holograms, the magnitude of signal remains high after cleaning of in-the-field particles. This is due to the signature of out-of-field particles which are close to the borders.
<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Accuracy (RMS error, in $\mu$m)</th>
<th>Running Time (in seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$x$</td>
<td>$y$</td>
</tr>
<tr>
<td>single-scale</td>
<td>0.054</td>
<td>0.040</td>
</tr>
<tr>
<td>($k_{max} = 0$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>multi-scale</td>
<td>0.053</td>
<td>0.041</td>
</tr>
<tr>
<td>($k_{max} = 2$)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Accuracy and computational time when going from a single scale to a pyramid with 3 scales. Accuracies are computed based on numerical simulations and are thus higher than what would be achievable on real data with imperfect modeling of the setup (see Sec. 4.A). Accuracies are comparable in both cases while the multi-scale method was about 4 times faster. We give the time required for the last (fine scale) parameter refinement compared to the first steps (either a single exhaustive search on the finest scale when $k_{max} = 0$ or an exhaustive search on the coarsest scale followed by successive refinements on finer scales).
5. Discussion and Conclusion

The inverse problems approach introduced in [19, 22, 31] is the optimal digital hologram reconstruction method in terms of accuracy. To mitigate its time costs, we have introduced a multi-scale algorithm which preserves the optimality of the inverse problems approach over classical approaches (e.g., hologram diffraction based methods). The main feature of this algorithm is to replace the computationally intensive exhaustive search by a coarse-to-fine processing. We analyzed the maximum down-sampling that is possible while avoiding erroneous particle localization. We have validated our algorithm using a collection of 100 simulated holograms and real holograms. The results indicate a factor of four increase in speed for a three layers multi-scale pyramid. This improvement makes it feasible to use inverse problems approaches to track parametric objects in videos of holograms (e.g., spherical particles time-resolved tracking in 3D with radius monitoring).

Choosing the sampling step sizes along $z$ and $r$ dimensions for the exhaustive search is often done in an over-conservative manner. The solution followed in this study was to use the shape of the cost function. Assuming that the size of the main basin of the cost function and its curvature change accordingly, one can estimate the minimum basin size from the value of parameters leading to the narrowest cost function (i.e., smallest CRLB). We chose the sampling step size such that the smallest basin gets 3 samples.

The time gain achieved by the multi-scale algorithm depends on the application. The algorithmic complexity of FAST is $\mathcal{O}(N/T_{k_{\text{max}}}^{2}(\log_{2}N - 2\log_{2}(T_{k_{\text{max}}})) \cdot n_{z} \cdot n_{r} + C \cdot N)$, number of samples in $z$ and $r$ dimensions respectively and $C$ a constant of order $\log_{2}(N)$. Computation of down-sampled versions of the hologram is negligible compared to the other steps of the algorithm (averaging followed by down-sampling is done in $\mathcal{O}(N)$). Thus for a wide range of $z$ and/or $r$ (i.e., larger parameter search space), the time cost of exhaustive step is the most significant one resulting in higher time gain of algorithm FAST.

The multi-scale algorithm results in faster particle detection and estimation only for particles located in the field of view of the camera, for the out-of-field particles’ signatures contain only high frequencies which would be filtered out during the down-sampling step. This limits the use of the multi-scale algorithm to in-the-field detection.

In this study, we considered only average filters to build the multi-resolution pyramid for two reasons: (i) after down-sampling, the noise is left uncorrelated; (ii) the expression of average filters in direct and Fourier domain are simple and fast to compute. Other low-pass filters with smaller frequency support (e.g., B-splines [41]) could also be considered, provided that adequate changes to the modeling (and consequently the calculation of CRLBs) are done and conditions (i) and (ii) are fulfilled. A trade-off must be found between the noise reduction, the aliasing prevention, and the preservation of high frequency content (below Nyquist frequency) for an accurate estimation. Determination of an optimal filter for the
pyramid construction is left as further work.

Thanks to the successive refinement steps, the proposed coarse-to-fine approach provides an early estimation of parameters with additional accuracy after each refinement step. These coarse results can provide a quick feedback for huge stacks of holograms generated by high-speed cameras while off-line processes can refine the estimations using the finer scales.

**Appendix A**

In this appendix we introduce an upper bound on the maximum period of down-sampling ($T_{k_{\text{max}}}$) and height of pyramid ($k_{\text{max}}$) using criteria on the number of preserved fringes on the down-sampled hologram. The low-pass filtering which is performed before down-sampling acts as a multiplication of two cardinal sine functions (on lateral dimensions X and Y) in the spatial domain. To keep $q$ fringes on the down-sampled hologram, the first zero of these cardinal sine functions should appear after $q$ maxima of the sine function of the model. The radial coordinate $\rho_q$ corresponding the the $q^{th}$ fringe is given by:

$$\pi \rho_q^2 = \frac{q}{\lambda z} 2\pi + \pi / 2$$

The x-coordinate $x_0$ of the first zero of the cardinal sine functions of filtering in the x direction is given by:

$$\frac{\pi x_0 T_k \kappa}{\lambda z} = \pi$$

where $\kappa$ is the pixel size. Therefore to have more than $q$ fringes at $x_0$, $\rho_q$ should be less than $x_0$:

$$\sqrt{\frac{2q + 1}{2}} \lambda z < \frac{\lambda z}{T_k}$$

Thus the maximum down-sampling period $T_{k_{\text{max}}}$ and height of the pyramid are:

$$T_{k_{\text{max}}} = \left\lfloor \frac{1}{\kappa} \sqrt{\frac{\lambda z_{\text{min}}}{2q + 1/2}} \right\rfloor, k_{\text{max}} = \lfloor \log_2 (T_{k_{\text{max}}}) \rfloor$$
where the brackets represent the floor function.

**Appendix B**

In this appendix we deduce a simplified formulation of the convolution of the Fresnel function with a real symmetric filter (e.g., object’s aperture). Considering \( h_z \) as the Fresnel function and \( f \) as a symmetric filter with bounded support, the convolution of this filter with the Fresnel function is given by:

\[
[h_z * f](x, y) = \frac{1}{i\lambda z} \int \int f(\xi, \eta) h_z(x - \xi, y - \eta) d\xi d\eta
\]

with

\[
h_z(x, y) = \frac{1}{i\lambda z} \exp\left(\frac{i\pi(x^2 + y^2)}{\lambda z}\right).
\]

Developing the phase of the Fresnel function gives

\[
[h_z * f](x, y) = \frac{1}{i\lambda z} \int \int f(\xi, \eta) \exp\left(\frac{i\pi}{\lambda z} \left[x^2 + \xi^2 + y^2 + \eta^2 - 2x\xi - 2y\eta\right]\right) d\xi d\eta.
\]

If the filter support satisfies the following

\[
\frac{\pi||\xi^2 + \eta^2||_{\text{max}}}{\lambda z} \ll \pi
\]

then:

\[
[h_z * f](x, y) \approx \frac{1}{i\lambda z} \exp\left(\frac{i\pi}{\lambda z} \left[x^2 + y^2\right]\right) \int \int f(\xi, \eta) \exp\left(-2\pi i \left[\frac{x}{\lambda z} \xi + \frac{y}{\lambda z} \eta\right]\right) d\xi d\eta
\]

which implies

\[
[h_z * f](x, y) \approx h_z(x, y) \cdot \mathcal{F}_{x \rightarrow \frac{x}{\lambda z}, y \rightarrow \frac{y}{\lambda z}}\{f\}
\]

with \( \mathcal{F}\{f\} \) representing the Fourier transform of \( f \). According to Eq. 17 the convolution of the Fresnel function with a real symmetric filter is simplified to the inner product of the Fresnel function and the Fourier transform of the filter. The filter could be objects’ apertures or the sensitive area of the pixel to model the pixel integration. Consequently
Eq. 17 can be used to simplify the analytical hologram model.

**Appendix C**

Simplification of Appendix B forces an upper bound on the size of filter $f$. It can be reformulated considering maximum aperture of objects (maximum radius) as $r_{\text{max}}$ and maximum width of filter as $T$. The resulting filter has the length $r_{\text{max}} + T\kappa/2$ on each direction where $\kappa$ is the pixel width (pixels are assumed to be square on the sensor). To ensure the satisfaction of $<<$ in Eq. 16 the following inequality should stand:

$$\frac{10\pi(r_{\text{max}} + T\kappa/2)^2}{\lambda z_{\text{min}}} < \pi$$

(18)

To have an order of magnitude for $z_{\text{min}}$ we take minimum $z$ such that the diffraction patterns obey the Nyquist theorem [11]:

$$z_{\text{min}} = z_{\text{Nyquist}} = \frac{L\kappa^2}{\lambda}$$

(19)

where $L$ is the width of the sensor in pixels.

Finally using Eq. 18 and Eq. 19 an upper bound on the size of the filter in pixels could be found as:

$$T < 2(\sqrt{\frac{L}{10}} - \frac{r_{\text{max}}}{\kappa})$$

Thus the maximum down-sampling period $T_{k_{\text{max}}}$ and height of the pyramid $k_{\text{max}}$ are:

$$T_{k_{\text{max}}} = \left\lfloor 2 \left(\sqrt{\frac{L}{10}} - \frac{r_{\text{max}}}{\kappa}\right) \right\rfloor, k_{\text{max}} = \left\lfloor \log_2(T) \right\rfloor$$

where the brackets represent the floor function.

**References**


34. L. Denis, D. A. Lorenz, and D. Trede, “Greedy solution of ill-posed problems: error