Inverse problems approaches for digital hologram reconstruction

Corinne Fournier[a,c], Loic Denis[b], Eric Thiebaut[b], Thierry Fournel[a], Mozhdeh Seifi[a]

[a]Laboratoire Hubert Curien, CNRS UMR 5516, Université Jean Monnet, 18 rue du Pr B. Lauras, Saint-Etienne, France;
[b]Centre de Recherche d'Astrophysique de Lyon, CNRS UMR 5574, Observatoire de Lyon, 9 avenue Charles André, Saint-Genis Laval Cedex, F-69561, France; Université Lyon 1, Villeurbanne, F-69622, France; Université de Lyon, Lyon, F-69000, France; and Ecole Normale Supérieure de Lyon, F-69000 Lyon, France;
[c]Telecom Saint-Etienne, F-42000 Saint-Etienne, France;

ABSTRACT

Digital holography (DH) is being increasingly used for its time-resolved three-dimensional (3-D) imaging capabilities. A 3-D volume can be numerically reconstructed from a single 2-D hologram. Applications of DH range from experimental mechanics, biology, and fluid dynamics. Improvement and characterization of the 3-D reconstruction algorithms is a current issue. Over the past decade, numerous algorithms for the analysis of holograms have been proposed. They are mostly based on a common approach to hologram processing: digital reconstruction based on the simulation of hologram diffraction. They suffer from artifacts intrinsic to holography: twin-image contamination of the reconstructed images, image distortions for objects located close to the hologram borders. The analysis of the reconstructed planes is therefore limited by these defects. In contrast to this approach, the inverse problems perspective does not transform the hologram but performs object detection and location by matching a model of the hologram. Information is thus extracted from the hologram in an optimal way, leading to two essential results: an improvement of the axial accuracy and the capability to extend the reconstructed field beyond the physical limit of the sensor size (out-of-field reconstruction). These improvements come at the cost of an increase of the computational load compared to (typically non iterative) classical approaches.

Keywords: Digital holography, inverse problems, image reconstruction techniques

1. INTRODUCTION

Digital Holography (DH) is a 3-D imaging technique which has been widely developed during the past few decades thanks to the enormous advances in digital imaging and computer technology. This technique achieves 3-D reconstruction of objects from a 2D hologram-image and reaches accuracies in the range of — or smaller than — the wavelength.[1–3] As 3-D information coded in a digital hologram can be recorded in one shot, this technique can be used with high speed cameras to perform time-resolved 3-D reconstructions of high speed phenomena.

DH is used in two types of applications: (i) the 3-D reconstruction of object surfaces (or optical index); (ii) the 3-D localization of micro-objects spread throughout a volume. Digital holography applications range from fluid flow measurement and structural analysis to medical imaging.[4–7] Off-axis setups are typically used in problems of type(i), while on-axis setups are best suited to problems of type (ii). Though the inverse problem methodology applies to both types of setups, we will focus in this paper on the latter.

Over the past decade, numerous algorithms for the analysis of digital holograms have been proposed (several journal special issues were published on the subject[2,8,9]). These algorithms are mostly based on a common approach to hologram processing (hereafter denoted as the classical approach): digital reconstruction based on the simulation of hologram diffraction. We recall in section 3 some artifacts that appear when such an approach

Further author information: (Send correspondence to corinne.fournier@univ-st-etienne.fr)

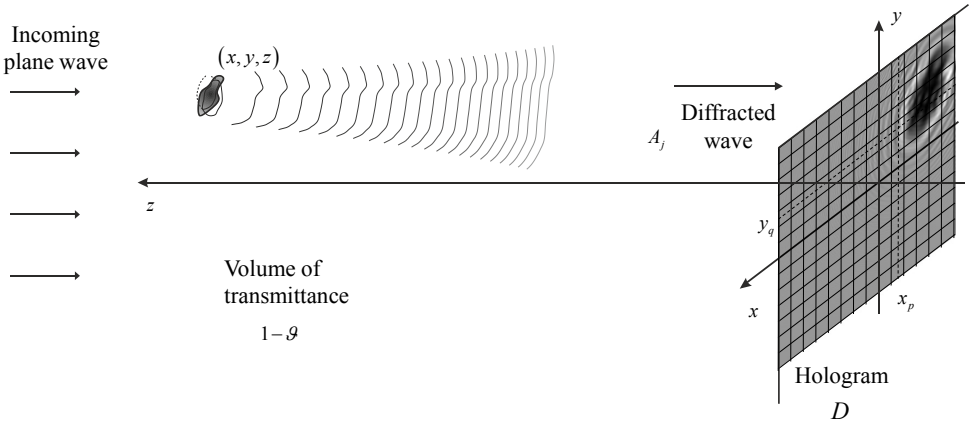


Figure 1: Illustration of in-line hologram formation model.

is used. To obtain a robust reconstruction method, these distortions must be either corrected *a posteriori* or are neglected, leading to sub-optimal techniques.

In contrast to this approach, the Inverse Problems (IP) perspective does not transform the hologram but rather search for the reconstruction that best models the measured hologram. This technique extracts more information from the hologram and solves two essential problems in digital holography: the improvement of the axial localization accuracy of an object and the enlargement of the studied field beyond the physical limit of the sensor size.[10] The drawback of this approach is a computational load heavier than that of the classical techniques.

We discuss in this paper the use of an IP approach for digital hologram reconstruction. We show that, depending on the application, different reconstruction algorithms can be derived. Special attention is paid to the problem of detection and location of micro-objects, a problem which is essential for many applications fields. Based on statistical estimation theory, we provide resolution bounds.

2. MODEL OF THE HOLOGRAM IMAGE FORMATION

In this section, we present the model of holograms[11] that will be used in the reconstruction methods. We consider an on-axis holography setup where studied objects are illuminated with a collimated laser beam, and the digital camera records both the object wave — diffracted light — and the reference wave — illuminating light — (see Figure 1). Considering n small objects of aperture ϑ_j and 3-D locations (x_j, y_j, z_j), the intensity I at the coordinate (x, y) on the hologram is given by

$$I(x,y) = I_{\text{src}} + I_{\text{bg}} - 2I_{\text{src}} \sum_{j=1}^{n} \eta_j \, \Re\big(A_j(x,y)\big) + I_{\text{src}} \sum_{i=1}^{n} \sum_{j=1}^{n} \eta_i \, A_i(x,y) \, \eta_j \, A_j^{\star}(x,y) \tag{1}$$

where I_{src} is the image level due to the laser source, I_{bg} is the background level, the real factors $\eta_j \in [0,1]$ account for possible variation of incident energy seen by an object due to non-uniform laser illumination and A_j is the term for amplitude diffraction by j^{th} object:

$$A_j(x,y) = \iint h_z(x-u, y-v)\vartheta_j(u,v) \, \mathrm{d}u \, \mathrm{d}v \tag{2}$$

where ϑ_j is the aperture function of the j^{th} object (a 2-D function — possibly complex-valued — defining the opacity of the object over a plane parallel to the hologram) and h_z is the impulse response function for free space

propagation (e.g. Fresnel function if Fresnel approximation is used) over a distance z (distance of the j^{th} object to the hologram). Equation 2 expresses a convolution.

For small objects the second order terms of equation (1) are negligible. The model then simplifies to a linear model:

$$I(x,y) = I_0 - \sum_{j=1}^{n} \alpha_j . m_j(x,y), \tag{3}$$

$$\text{with} \quad m_j(x,y) = \iint h_{z_j}(x-u, y-v)\mathcal{O}_j(u,v) \equiv \left[h_{z_j} * \mathcal{O}_j\right](x,y),$$

($\mathcal{O}_j$ stands for the aperture of the j^{th} object)

$$I_0 = I_{\text{src}} + I_{\text{bg}},$$

$$\alpha_j = 2\, I_{\text{src}}\, \eta_j .$$

The digitization of intensity I on an N^2 pixels camera leads to a hologram D — or a vector $\boldsymbol{d}$ of N^2 grayvalues in matrix notation — which may be related to the diffraction pattern of each object (FI), or to the opacity distribution of the objects (FII):

$$\text{(FI)} \qquad \begin{bmatrix} D(x_1,y_1) \\ \vdots \\ D(x_N,y_N) \end{bmatrix} = \begin{bmatrix} I_0 - \sum_j \alpha_j m_j(x_1,y_1) + \epsilon_1 \\ \vdots \\ I_0 - \sum_j \alpha_j m_j(x_1,y_1) + \epsilon_{N^2} \end{bmatrix} \qquad \leftrightarrow \qquad \boldsymbol{d} = I_0 \cdot \mathbf{1} - \boldsymbol{M}\,\boldsymbol{\alpha} + \boldsymbol{\epsilon} \tag{4}$$

$$\text{(FII)} \qquad \begin{bmatrix} D(x_1,y_1) \\ \vdots \\ D(x_N,y_N) \end{bmatrix} = \begin{bmatrix} I_0 - \sum_k \left[h_{z_k} * \vartheta_k\right](x_1,y_1) + \epsilon_1 \\ \vdots \\ I_0 - \sum_k \left[h_{z_k} * \vartheta_k\right](x_N,y_N) + \epsilon_{N^2} \end{bmatrix} \qquad \leftrightarrow \qquad \boldsymbol{d} = I_0 \cdot \mathbf{1} - \boldsymbol{H}\,\boldsymbol{\vartheta} + \boldsymbol{\epsilon} \tag{5}$$

Equations (4) and (5) are written in compact form in matrix notation. In words, equation (4) expresses the recorded hologram $\boldsymbol{d}$ as the sum of a constant offset ($I_0 \cdot \mathbf{1}$), the diffraction patterns of each object ($\boldsymbol{M}\,\boldsymbol{\alpha}$) and a perturbation term accounting for the different sources of noise and for our modeling approximations ($\boldsymbol{\epsilon}$). The term $\boldsymbol{M}\,\boldsymbol{\alpha}$ is the product between a $N^2 \times n$ matrix ($\boldsymbol{M}$) and a n elements vector ($\boldsymbol{\alpha}$). Matrix $\boldsymbol{M}$ may be thought of as a dictionary of the diffraction pattern of each of the n objects (the j-th column of matrix $\boldsymbol{M}$ corresponds to the N^2 graylevels of the diffraction pattern of the j-th object: $[m_j(x_1,y_1), \cdots, m_j(x_N,y_N), \cdots, m_j(x_N,y_N)]^t$). Vector $\boldsymbol{\alpha}$ defines the amplitude of each of the n diffraction patterns. Equation (4) thus corresponds to a discretization of equation (3).

Equation (5) expresses the hologram $\boldsymbol{d}$ as the sum of an offset ($I_0 \cdot \mathbf{1}$), the diffraction patterns $\boldsymbol{H}\,\boldsymbol{\vartheta}$ created by the opacity distribution $\boldsymbol{\vartheta}$, and a noise term ($\boldsymbol{\epsilon}$). If the opacity distribution is defined over K planes of L^2 pixels, $\boldsymbol{\vartheta}$ is a vector of $K \cdot L^2$ elements corresponding to the stacking of all opacity values. $\boldsymbol{H}$ is then a $N^2 \times K \cdot L^2$ matrix corresponding to a (discrete) diffraction operator. Each column of $\boldsymbol{H}$ is a discretization of the impulse response kernel h, i.e., the diffraction pattern on the hologram created by a point-like opaque object at a given 3-D location. $\boldsymbol{H}\,\boldsymbol{\vartheta}$ corresponds to the summation of the convolution of the opacity distribution in each plane z by the impulse response kernel of distance z.

Matrices $\boldsymbol{M}$ and $\boldsymbol{H}$ are written formally to clarify the proposed models and the derived reconstruction in the subsequent sections. It is worth noting that, in practice, they are neither stored nor explicitly multiplied to vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\vartheta}$. Due to (transversal) shift-invariance of models m_j and kernels h_j, the products $\boldsymbol{M}\,\boldsymbol{\alpha}$ and $\boldsymbol{H}\,\boldsymbol{\vartheta}$ can be computed using few fast Fourier transforms.

Pixel integration on the camera can be taken into account in matrices $\boldsymbol{M}$ and $\boldsymbol{H}$ by convolving the diffraction patterns m_j and diffraction kernels h_{z_k} (which form the matrix columns) with the pixel's sensitive area.

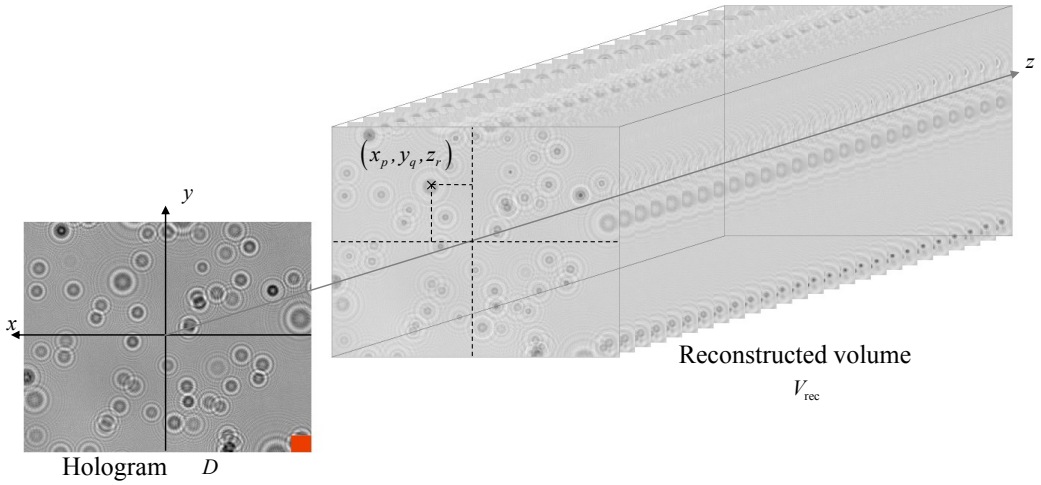


Figure 2: Illustration of classical reconstruction based on hologram diffraction. The z axis is magnified versus x and y axis. The red rectangle corresponds to the real size of the hologram.

3. DIGITAL HOLOGRAM RECONSTRUCTION BASED ON HOLOGRAM DIFFRACTION

The large majority of methods for reconstructing digital holograms are based on the simulation of an optical reconstruction, followed by a 3-D analysis of the reconstructed volume. In all-optical holography, after a hologram has been recorded and the holographic plate has been processed, the plate is re-illuminated with the reference wave. Hologram diffraction creates a virtual (i.e., defocused) and a real (i.e., focused) image. In digital holography, the holographic plate is replaced by a digital camera whose sensor size and definition is worse by several orders of magnitude. The simulation of hologram diffraction, though straightforward to implement (and fast), lead to sub-optimal reconstructions with distortions due to boundary effects and the presence of the virtual (twin) image.

In this section, we summarize hologram-diffraction based approaches used in DH and their limitations. We then present, in more details, some of their drawbacks and point out some of the works in the literature that describe them.

3.1 Classical reconstruction of digital holograms

The classical 3-D reconstruction of digital holograms is performed in two steps. The first step is based on a numerical simulation of the optical reconstruction. A 3-D image volume V_{rec} is obtained by computing the diffracted field in planes located at increasing distances from the hologram (see Figure 2). Different techniques to simulate diffraction have been proposed (Fresnel transform,[5] fractional Fourier transform,[12,13] wavelets transform[14,15]). Using a convolution-based diffraction model, V_{rec} is given by:

$$V_{\text{rec}}(x_p, y_q, z_r) = [D * h_{z_r}](x_p, y_q) \quad \leftrightarrow \quad \boldsymbol{v} = \boldsymbol{H}^t \boldsymbol{d} \tag{6}$$

Unfortunately, hologram diffraction does not invert hologram recording: operator $\boldsymbol{H}^t\boldsymbol{H}$ is far from the identity (i.e., the impulse response of the system "hologram recording" + "linear reconstruction" is a spatially variant halo with large spread along the axial direction z).

The second step consists of localizing and sizing each object in the obtained 3-D image. The best focusing plane for each object has to be detected. Various criteria are suggested in the literature. Some are based on the local analysis of the sampled reconstructed volume. For example, one searches for the minimum of the gray level on the z-axis crossing the object center[16] or computes the barycenter of the labeled object image after thresholding of the 3-D reconstructed image.[17] Some authors use the imaginary part of the reconstructed field.[18] Other approaches are based on an analysis of the object's 3-D image. Liebling uses the criterion of the

sparsity of wavelets coefficients[19] and Dubois uses the minimization of the integrated reconstructed amplitude.[20] Hologram-diffraction based approaches suffer from various limitations:

- the lateral field of view is limited and, in practice, must be restricted to the center of the reconstructed images to reduce the border effects;
- under-sampled holograms can lead to artifacts (e.g. ghost images);
- spurious twin-images of the objects get reconstructed;
- multiple focusing can occur around the actual depth location of each object;[21]

3.2 Limits of classical reconstruction

This section is aimed at detailing the first two limitations of classical reconstruction listed above. The next section explains how the IP approach push back these limits.

Due to technological constraints, digital holography suffers from the bad resolution of digital cameras (about 50 times worse than holographic plates). For a correct sampling of the digital hologram, the maximal spatial frequency in the image is imposed by the Nyquist criterion and thus by the pixel sampling (p). When Nyquist criterion is not fulfilled (signal frequency higher than Nyquist frequency) an aliasing phenomenon appears and produces artifacts or ghost images in the reconstructed images.[22–25] These artifacts can lead to false object detections. Some solutions have been proposed to overcome this problem. Onural and Stern[23,25] suggest a filtering of these ghost images in the reconstructed planes (in the case of known location of the true object). Jacquot[22] presents an over-sampling of the hologram and Coupland[24] suggests removing ghost images by using an irregular sampling of the signal, which involves a decrease of the amplitude of these images compared to the real ones. Let us notice that these last two methods require some heavier experimental setups.

To avoid ghost images occurrence in the hologram, Nyquist criterion should be respected. For an object located on the optical axis at a distance z from the hologram, the following relation must be satisfied[5] $z > \frac{Lp}{\lambda}$, where L is the width of the sensor and λ is the laser wavelength. To verify this condition, either a high resolution sensor has to be used (small $\Delta\xi$) or the object has to be at a minimum distance $z_{min} = \frac{Lp}{\lambda}$, where L stand for the width of the sensor considered squared. The camera cannot therefore be laid-out too close from the objects. This limits the contrast of the interference fringes (the contrast decreases when the distance camera-objects z increases) and limit the numerical aperture, and thus the resolution.

An other drawback which can prove embarrassing within the framework of metrology is the appearance of some artifacts in the reconstructed 3-D image due to hologram truncation. This problem is all the more important since digital camera sensors are small compared to holographic plates (their surface is hundred times less). The loss of the fringes outside of the sensor is at the origin of artifacts named "border effects". This phenomenon distorts the reconstructed object and generates errors in the 3-D location and size estimation. This phenomenon is illustrated on figure 3. The figure shows the distortions of a particle located on the border of an experimental hologram of water droplets on several consecutive reconstructed planes.

Hologram-diffraction based techniques require an expansion of the hologram outside its boundaries. The hologram can either be zero-padded, implicitly periodized by use of discrete Fourier transform. In each case, the fringes recorded are completed by erroneous values outside the hologram, which leads to distorted low-contrast 3-D images (and possibly ghost objects). The contrast of these reconstructed images is very faint and the artifacts are very strong (figure 3). These artifacts, generating a loss of accuracy in the 3-D reconstruction, restrict the usability of these reconstruction approaches to the area located at the hologram center (where interference patterns truncation is weak). Several methods have been proposed to reduce the border effects. Dubois suggested a technique which extrapolates the hologram.[26] Cuche proposed a method based on a weighting of the image border by a cubic spline therefore reducing border effects at the cost of a loss of contrast close to the hologram borders.[27] We show next section that using an inverse problem approach not only solves these border effects, but also makes it possible to reconstruct out-of-field objects.

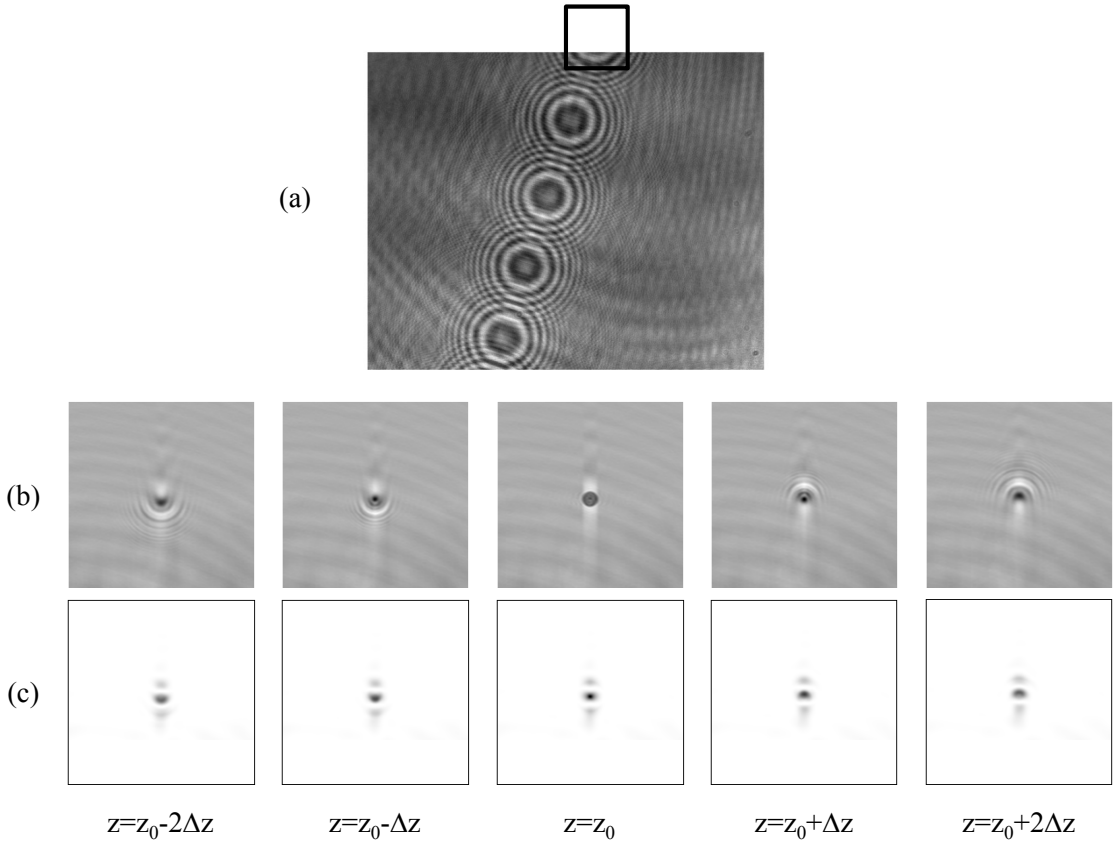


Figure 3: Illustration of classical reconstruction compared with the criterion map : (a) experimental in-line hologram of droplets (b) classical reconstruction based on hologram diffraction at different depths z. Artefacts appear during the numerical reconstruction due to the truncation of diffraction rings on the hologram boundary. (c) Criterion computation based on IP at different depths z (see eq. 21). The images represented in (b) and (c) correspond to the square area drawn on the hologram (a).$z_0 = 0.273m$ corresponds to the in-focus distance, $\Delta z = 6mm$.

4. OPTIMAL HOLOGRAM PROCESSING: THE INVERSE PROBLEM APPROACH

We described in section 2 two linear models of a hologram: equations (4) and (5). The amplitude of the objects $\boldsymbol{\alpha}$ or the opacity distribution $\boldsymbol{\vartheta}$ can be estimated by inverting the hologram formation models, using a suitable regularization: this corresponds to the inverse problem approach.

The noise term $\boldsymbol{\epsilon}$ in the hologram models can be considered Gaussian, with an inverse covariance matrix $\boldsymbol{W}$. Data are then distributed following a distribution of the form:

$$\text{(FI)} \quad \mathrm{p}(\boldsymbol{d}|\boldsymbol{\alpha}) \propto \exp\left[-(I_0 \cdot \mathbf{1} - \boldsymbol{M}\,\boldsymbol{\alpha} - \boldsymbol{d})^t \boldsymbol{W} (I_0 \cdot \mathbf{1} - \boldsymbol{M}\,\boldsymbol{\alpha} - \boldsymbol{d})\right] \tag{7}$$

$$\text{(FII)} \quad \mathrm{p}(\boldsymbol{d}|\boldsymbol{\vartheta}) \propto \exp\left[-(I_0 \cdot \mathbf{1} - \boldsymbol{H}\,\boldsymbol{\vartheta} - \boldsymbol{d})^t \boldsymbol{W} (I_0 \cdot \mathbf{1} - \boldsymbol{H}\,\boldsymbol{\vartheta} - \boldsymbol{d})\right] \tag{8}$$

Noise is generally considered white, so that $\boldsymbol{W}$ is diagonal: $\boldsymbol{W} = \mathrm{diag}(\boldsymbol{w})$. Non uniform $\boldsymbol{w}$ can account for a signal-dependant variance or can be used to restrict the support to the hologram boundaries (i.e., $w_k = 0$ for pixels k that are outside the hologram support).

To simplify the expressions in the following derivations, we define the scalar product $<\boldsymbol{u}, \boldsymbol{v}>_{\boldsymbol{W}}$ and the induced norm $\|\boldsymbol{u}\|^2_{\boldsymbol{W}}$ as follows:

$$<\boldsymbol{u}, \boldsymbol{v}>_{\boldsymbol{W}} \;=\; \frac{\boldsymbol{u}^t \boldsymbol{W} \boldsymbol{v}}{\mathbf{1}^t \boldsymbol{W} \mathbf{1}} \quad \left(= \frac{\sum_k w_k u_k v_k}{\sum_k w_k} \text{ for a diagonal } \boldsymbol{W}\right) \tag{9}$$

$$\|\boldsymbol{u}\|^2_{\boldsymbol{W}} \;=\; <\boldsymbol{u}, \boldsymbol{u}>_{\boldsymbol{W}} \;=\; \frac{\boldsymbol{u}^t \boldsymbol{W} \boldsymbol{u}}{\mathbf{1}^t \boldsymbol{W} \mathbf{1}} \quad \left(= \frac{\sum_k w_k u_k^2}{\sum_k w_k} \text{ for a diagonal } \boldsymbol{W}\right) \tag{10}$$

The negative log-likelihood is given, up to an additive and a multiplicative constant, by:

$$\text{(FI)} \quad -\log \mathrm{p}(\boldsymbol{d}|\boldsymbol{\alpha}) = \|I_0 \cdot \mathbf{1} - \boldsymbol{M}\,\boldsymbol{\alpha} - \boldsymbol{d}\|^2_W \tag{11}$$

$$\text{(FII)} \quad -\log \mathrm{p}(\boldsymbol{d}|\boldsymbol{\vartheta}) = \|I_0 \cdot \mathbf{1} - \boldsymbol{H}\,\boldsymbol{\vartheta} - \boldsymbol{d}\|^2_W \tag{12}$$

Maximum likelihood estimation of the offset: The offset that maximizes the likelihood is:

$$\text{(FI)} \quad I_0^\dagger = <\mathbf{1}, \boldsymbol{M}\,\boldsymbol{\alpha} + \boldsymbol{d}>_{\boldsymbol{W}} \tag{13}$$

$$\text{(FII)} \quad I_0^\dagger = <\mathbf{1}, \boldsymbol{H}\,\boldsymbol{\vartheta} + \boldsymbol{d}>_{\boldsymbol{W}} \tag{14}$$

The neg-log-likelihood $\mathcal{L}$ can then be re-written in the simplified form:

$$\text{(FI)} \quad \mathcal{L}_{\mathrm{I}}(\boldsymbol{d}, \boldsymbol{\alpha}) = -\log \mathrm{p}(\boldsymbol{d}|\boldsymbol{\alpha}) = \|\bar{\boldsymbol{M}}\,\boldsymbol{\alpha} - \bar{\boldsymbol{d}}\|^2_W \tag{15}$$

$$\text{(FII)} \quad \mathcal{L}_{\mathrm{II}}(\boldsymbol{d}, \boldsymbol{\vartheta}) = -\log \mathrm{p}(\boldsymbol{d}|\boldsymbol{\vartheta}) = \|\bar{\boldsymbol{H}}\,\boldsymbol{\vartheta} - \boldsymbol{d}\|^2_W \tag{16}$$

with the centered variables:

$$\bar{\boldsymbol{d}} = \boldsymbol{d} - \mathbf{1}<\mathbf{1}, \boldsymbol{d}>_{\boldsymbol{W}} \quad \left(= \begin{bmatrix} d_1 - \sum_k w_k d_k / \sum_k w_k \\ \vdots \\ d_{N^2} - \sum_k w_k d_k / \sum_k w_k \end{bmatrix} \text{ for a diagonal } \boldsymbol{W}\right)$$

$$\bar{\boldsymbol{M}} = [\bar{\boldsymbol{m}}_1, \ldots, \bar{\boldsymbol{m}}_n]$$

$$\forall j,\ \bar{\boldsymbol{m}}_j = \mathbf{1}<\mathbf{1}, \boldsymbol{m}_j>_{\boldsymbol{W}} - \boldsymbol{m}_j \quad \left(= \begin{bmatrix} \sum_k w_k m_k / \sum_k w_k \ - m_1 \\ \vdots \\ \sum_k w_k m_k / \sum_k w_k \ - m_{N^2} \end{bmatrix} \text{ for a diagonal } \boldsymbol{W}\right)$$

$$\bar{\boldsymbol{H}} = [\bar{\boldsymbol{h}}_{z_1}, \ldots, \bar{\boldsymbol{h}}_{z_K}]$$

$$\forall k,\ \bar{\boldsymbol{h}}_k = \mathbf{1}<\mathbf{1}, \boldsymbol{h}_k>_{\boldsymbol{W}} - \boldsymbol{h}_k \quad \left(= \begin{bmatrix} \sum_k w_k h_k / \sum_k w_k \ - h_1 \\ \vdots \\ \sum_k w_k h_k / \sum_k w_k \ - h_{N^2} \end{bmatrix} \text{ for a diagonal } \boldsymbol{W}\right)$$

When the objects can be parameterized (e.g., disks), the objects can be detected and localized by using form (FI), as detailed in section 4.1. More complex objects require the reconstruction of the opacity distribution using form (FII), see section 4.2.

4.1 Detection and location of parameterized objects

When studying objects that can be parameterized with few parameters (3-D location and shape), a dictionary $\boldsymbol{M}$ of diffraction patterns can be considered to model the hologram (form FI). Since the 3-D location of an object is continuous, the dictionary $\boldsymbol{M}$ should also be continuous (i.e., with infinitely many elements). The approach proposed in[28,29] solves the problem in two steps:

- a global detection step, which finds the best-matching element in a discrete dictionary $\boldsymbol{M}$(i.e., the diffraction pattern for a given 3-D location and shape),
- a local optimization step, which fits the selected diffraction pattern to the data for sub-pixel estimation.

The objects are detected one after the other, and each time an object has been detected and located (with sub-pixel accuracy) its contribution on the hologram is subtracted. The procedure is then repeated on the residuals. This approach for hologram reconstruction corresponds to the class of greedy algorithms[30] known in signal processing as Matching Pursuit,[31] or in radio-astronomy as CLEAN algorithm.[32]

Global object detection: In the first step, the best matching diffraction pattern of dictionary $\boldsymbol{M}$ is researched. The element that leads to the largest decrease of the neg-log-likelihood $\mathcal{L}_{\mathrm{I}}$ is identified as the most probable (i.e., detected):

$$\underset{\substack{\alpha\geq 0\\ \bar{\boldsymbol{m}}\in\{\bar{\boldsymbol{m}}_1,\ldots,\bar{\boldsymbol{m}}_n\}}}{\arg\min} \quad \|\alpha\,\bar{\boldsymbol{m}}-\bar{\boldsymbol{d}}\|_W^2 \tag{17}$$

The optimal amplitude $\alpha^\dagger$ for a given diffraction pattern $\bar{\boldsymbol{m}}$ is:

$$\alpha^\dagger(\bar{\boldsymbol{m}})=\frac{<\bar{\boldsymbol{m}},\bar{\boldsymbol{d}}>_{\boldsymbol{W}}}{\|\bar{\boldsymbol{m}}\|_{\boldsymbol{W}}^2} \quad \text{if} \quad <\bar{\boldsymbol{m}},\bar{\boldsymbol{d}}>_{\boldsymbol{W}}\geq 0, \quad \text{otherwise} \quad \alpha^\dagger=0. \tag{18}$$

By replacing α by its optimal value in equation (17), the diffraction pattern $\bar{\boldsymbol{m}}^\dagger$ that minimizes $\mathcal{L}_{\mathrm{I}}$ is given by:

$$\bar{\boldsymbol{m}}^\dagger = \underset{\bar{\boldsymbol{m}}\in\{\bar{\boldsymbol{m}}_1,\ldots,\bar{\boldsymbol{m}}_n\}}{\arg\min} \quad \mathscr{C}(\bar{\boldsymbol{m}}) \qquad \text{subject to} \qquad <\bar{\boldsymbol{m}},\bar{\boldsymbol{d}}>_{\boldsymbol{W}}\geq 0 \tag{19}$$

$$= \underset{\bar{\boldsymbol{m}}\in\{\bar{\boldsymbol{m}}_1,\ldots,\bar{\boldsymbol{m}}_n\}}{\arg\min} \quad \|\bar{\boldsymbol{m}}\|_{\boldsymbol{W}}^2\cdot\alpha^\dagger(\bar{\boldsymbol{m}})^2 \tag{20}$$

$$\text{with } \mathscr{C}(\bar{\boldsymbol{m}})=-\frac{<\bar{\boldsymbol{m}},\bar{\boldsymbol{d}}>_{\boldsymbol{W}}^2}{\|\bar{\boldsymbol{m}}\|_{\boldsymbol{W}}^2} \tag{21}$$

The object detected is the one whose diffraction pattern is the most correlated with the data: $\alpha^\dagger$ corresponds to a normalized correlation between a model and the hologram. Since the diffraction patterns are shift-invariant, it can be shown that the correlations in equations (18) and (19) can be computed using fast Fourier transforms.[29]

Note that this global detection compared with the classical reconstruction is less sensitive to ghost images.[10] Furthermore, "Border effects" which classically lead to measurement bias, are removed by taking into account the boundaries of the sensor by means of $\boldsymbol{W}$. Figure 3.c shows the values of the criterion $\mathscr{C}(\bar{\boldsymbol{m}})$ on several consecutive reconstructed planes. Unlike for classical reconstruction, the minimum criterion value in these planes is on the in-focus plane.

Local optimization: Once a diffraction pattern has been selected, its parameters (3-D location and shape) can be fitted to lead to sub-pixel accuracy.

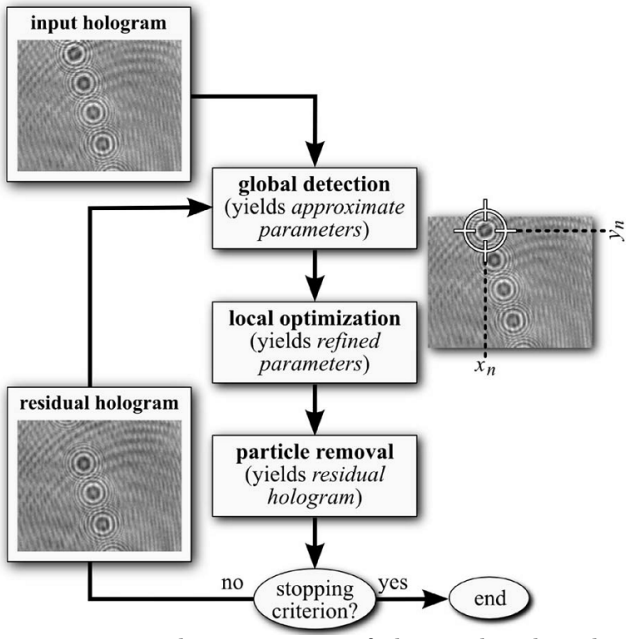


Figure 4: Iterative algorithm to estimate the parameters of objects distributed in a volume (Soulez *et al.*[29]).

Figure 4 and 5 illustrate an application of this algorithm to detect spherical opaque particles. Matrix $\boldsymbol{W}$ accounts for the hologram support and makes it possible to detect particles even when located far outside of the field of view (Fig. 5).

Accurate 3-D location of spherical objects by model fitting on the hologram as been applied to water droplets with diameters about $100\mu m$ by Soulez *et al.*[28,29] and to colloidal spherical particles about $1\mu m$ using DH microscopy by Grier *et al.*[33,34]

4.2 Reconstruction of 3-D transmittance distributions

When the objects are too complex to be parameterized by few parameters, or when the purpose is to reconstruct unknown objects, form (FII) is considered: a 3-D transmittance distribution is reconstructed from the hologram. Due to the ill-posed nature of this inversion problem, it is mandatory to regularize the problem. The reconstructed 3-D distribution $\boldsymbol{\vartheta}$ is then given by the maximum a posteriori estimate (MAP):

$$\boldsymbol{\vartheta}^{(\mathrm{MAP})} \quad = \quad \underset{\boldsymbol{\vartheta}}{\arg\min} \quad \|\bar{\boldsymbol{H}}\boldsymbol{\vartheta} - \bar{\boldsymbol{d}}\|_W^2 \; + \; \beta\, \Phi_{\mathrm{reg}}(\boldsymbol{\vartheta}) \tag{22}$$

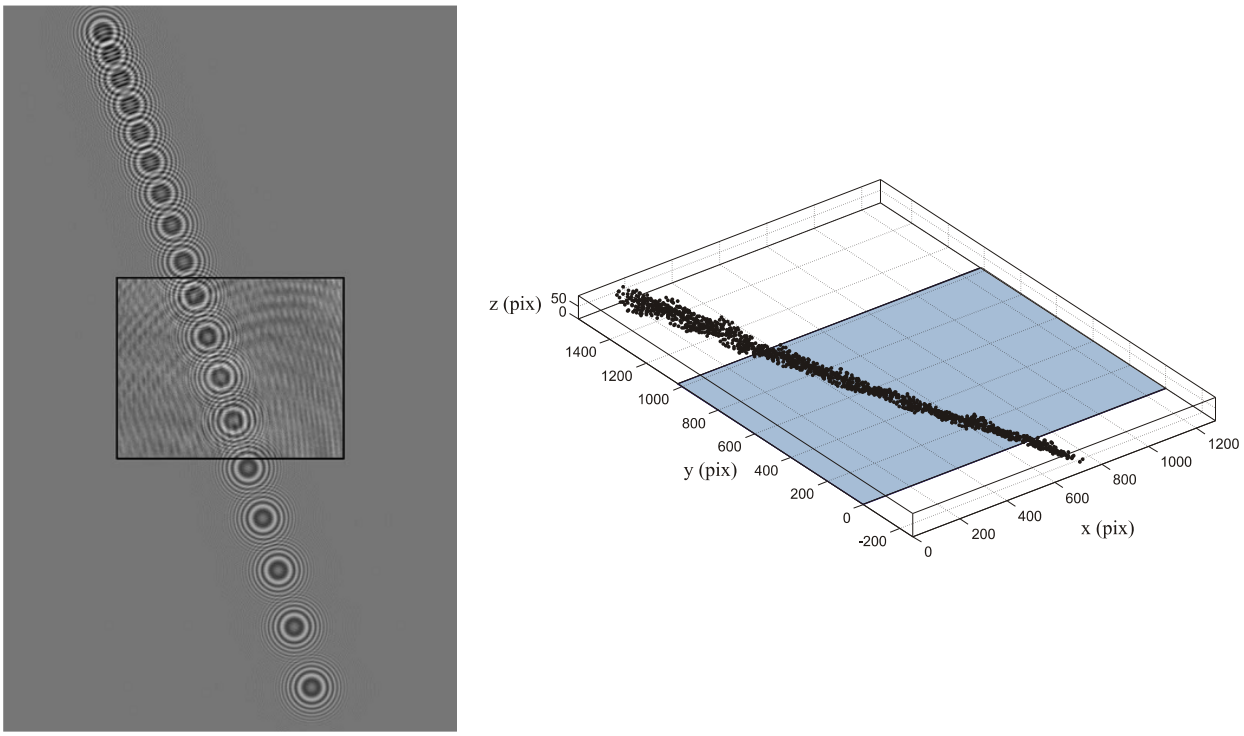
Several regularizations Φ_{reg} have been proposed to reconstruct holograms. When extend objects are considered, an edge-preserving smoothness prior like total variation (the sum of the spatial gradient norm) is generally chosen:[35–38]

$$\boldsymbol{\vartheta}^{(\mathrm{MAP})} \quad = \quad \underset{\boldsymbol{\vartheta}}{\arg\min} \quad \|\bar{\boldsymbol{H}}\boldsymbol{\vartheta} - \bar{\boldsymbol{d}}\|_W^2 \; + \; \beta\, \mathrm{TV}(\boldsymbol{\vartheta}), \qquad \text{with} \qquad \mathrm{TV}(\boldsymbol{\vartheta}) = \sum_k \sqrt{(\boldsymbol{D_x}\,\boldsymbol{\vartheta})_k^2 + (\boldsymbol{D_y}\,\boldsymbol{\vartheta})_k^2} \tag{23}$$

where $\boldsymbol{D_x}$ and $\boldsymbol{D_y}$ are the finite difference operators along x and y (i.e., tranversal) axes.

It has been shown[39] that enforcing a sparsity constraint through an ℓ^1 norm is sufficient to reconstruct holograms of diluted volumes:

$$\boldsymbol{\vartheta}^{(\mathrm{MAP})} \quad = \quad \underset{\boldsymbol{\vartheta}}{\arg\min} \quad \|\bar{\boldsymbol{H}}\boldsymbol{\vartheta} - \bar{\boldsymbol{d}}\|_W^2 \; + \; \beta\, \|\boldsymbol{\vartheta}\|_1, \qquad \text{with} \qquad \|\boldsymbol{\vartheta}\|_1 = \sum_k |\vartheta_k| \tag{24}$$



(a) (b)

Figure 5: Illustration of droplets detection located out-of-field (from Soulez *et al.*[29]): (a) superimposition of one hologram of the serie and the model of this hologram calculated from 16 detected particles (including 12 out-of-field); (b) represents 3D jet obtained by the detection of all particles located in a field corresponding to more than 16 times the hologram surface. The corresponding surface of the sensor is represented in blue. The droplets detection is realized without significant bias even for particles located far away from the sensor.

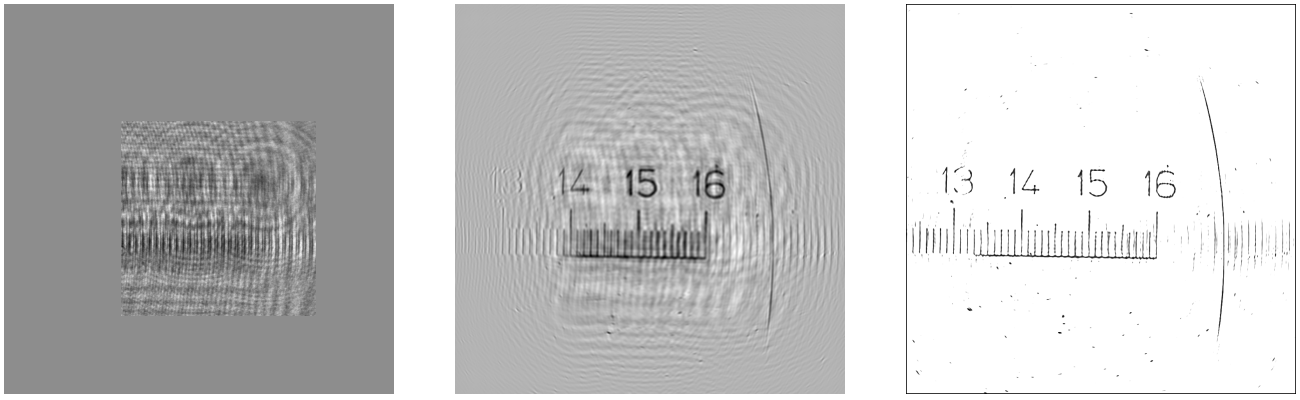


Figure 6: Reconstruction of an experimental Gabor hologram of target: hologram (left); classical linear reconstruction (center); MAP estimate with sparsity enducing prior of equation (24) and positivity constraint[39] (right). Regularized reconstruction of holograms makes it possible to extend the field the view and suppresses twin-image artifacts.

A positivity constraint and spatially-variant regularization weights $\Phi_{\text{reg}}(\boldsymbol{\vartheta}) = \sum_k \beta_k |\vartheta_k|$ improve the reconstruction and makes it possible to extend the field of view, as illustrated in Figure 6.

Note that the ℓ^1 norm minimization can also be applied to the object detection problem described in previous paragraph. Joint detection of all objects is more robust, in the case of many objects, than iterative detection of one object at a time. Intermediate procedures have been proposed in the compressed sensing literature[40] that detect several objects at a time, in a greedy fashion, and which can be adapted to include the local optimization step used to model a continuous dictionary.

5. ESTIMATION OF THE ACCURACY

The estimation and the improvement of the accuracy are key issues of DH.[41–44] As the accuracy depends on several experimental parameters (e.g., sensor definition, fill factor, and recording distance) experimenters are in need of methodologies to tune the experimental setup and to select the reconstruction that will provide the best achievable accuracy. The commonly used approach for accuracy estimation is to evaluate the Rayleigh resolution by estimating the width of the point spread function of the digital holographic system in the reconstructed planes.[41,43–45] We suggested recently[46] a methodology based on parametric estimation theory (see Kay[47]) to estimate the single point resolution[48](i.e., the standard deviation on the 3-D coordinates of a point source) in on-axis DH. This methodology can be applied to many DH configurations by adapting the hologram formation model, and possibly changing the noise model.

5.1 Cramér-Rao lower bound

According to Cramér-Rao inequality, the covariance matrix of any unbiased estimator $\hat{\boldsymbol{\theta}} = \{\hat{\theta}_i\}_{i=1:n_p}$ of the unknown vector parameter $\boldsymbol{\theta}^*$ is bounded from below by the inverse of the so-called Fisher information matrix:

$$var\left(\hat{\theta}_i\right) \geq \left[\boldsymbol{I}^{-1}\left(\boldsymbol{\theta}^*\right)\right]_{i,i} \tag{25}$$

where $\boldsymbol{I}\left(\boldsymbol{\theta}^*\right)$ is the $n_p \times n_p$ Fisher information matrix.

Fisher information matrix is defined from the gradients of the log-likelihood function $\log p(\mathbf{d};\boldsymbol{\theta})$:[47]

$$\left[\boldsymbol{I}\left(\boldsymbol{\theta}\right)\right]_{i,j} \stackrel{\text{def}}{=} E\left[\frac{\partial \log p(\mathbf{d};\boldsymbol{\theta})}{\partial \theta_i}\frac{\partial \log p(\mathbf{d};\boldsymbol{\theta})}{\partial \theta_j}\right]. \tag{26}$$

where $\boldsymbol{\theta}$ stands for the parameters vector of the object.

In the case of additive white Gaussian noise model (see section 4.1), Fisher information matrix can be computed using gradients of the model $\boldsymbol{m}(\boldsymbol{\theta})$:[46]

$$\left[\boldsymbol{I}(\boldsymbol{\theta})\right]_{i,j} = \alpha^2 \left\langle \frac{\partial \boldsymbol{m}(\boldsymbol{\theta})}{\partial \theta_i}, \frac{\partial \boldsymbol{m}(\boldsymbol{\theta})}{\partial \theta_j} \right\rangle_{\boldsymbol{W}} \tag{27}$$

Note that $\boldsymbol{W}$ accounts for the finite size of the sensor. The CRLB is asymptotically (for large samples) reached by maximum likelihood estimators. In digital holography, where the signal is distributed on the whole sensor, estimation is performed using a large set of independent identically distributed measurements (typically more than one million). The maximum likelihood estimator then approaches the Cramér Rao Lower Bound. Note that if the optimization technique used for maximization of the likelihood fails to reach the global minimum or if the noise level is too high, the resulting estimation error will exceed CRLB.

5.2 Single point resolution maps

A previous study[46] about single point resolution estimation lead to closed-form expressions of resolutions. It showed that:

- the CRLB predicted resolution behaves on optical axis as the classical Rayleigh resolution predicts;
- the resolution depends on the lateral coordinate of the point source;
- estimated parameters are correlated (an error on one parameter influences the estimation of the others) ;

Examples of standard deviation maps calculated using the described methodology are presented in Figure 7.

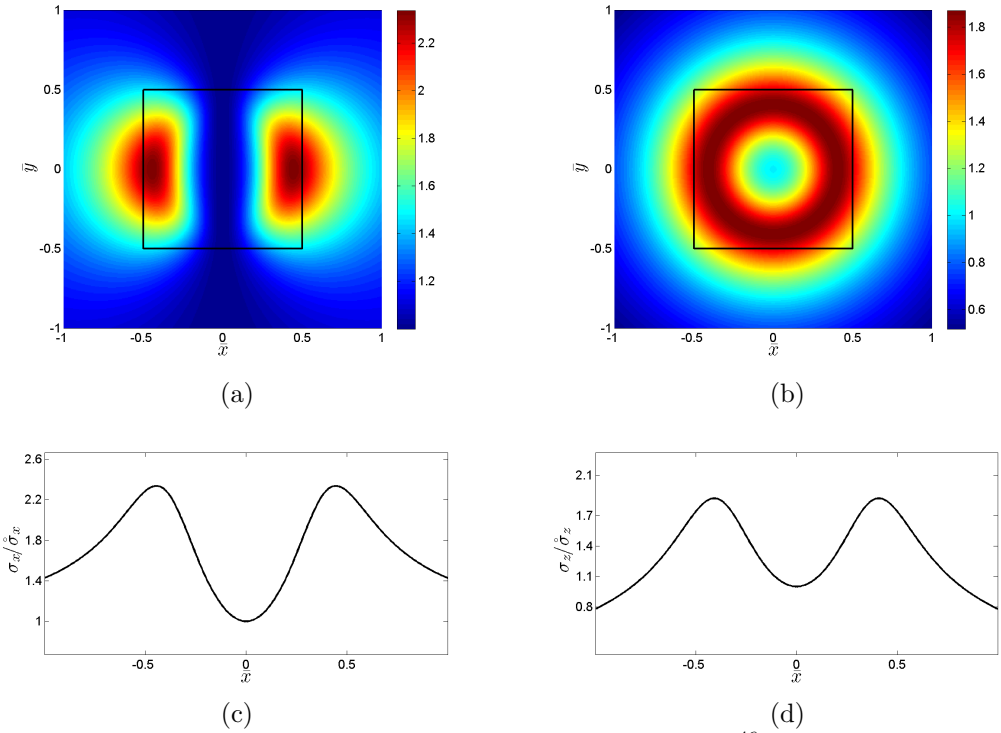


Figure 7: Single point resolution in a transversal plane (from Fournier *et al.*[46]): (a) x-resolution map normalized by the value of x-resolution on the optical axis; (b) normalized z-resolution map; (c) x-resolution for $\bar{y} = 0$; (d) z-resolution for $\bar{y} = 0$; for $z = 100mm$, $\lambda = 0.532\mu m$, $\Omega = 8.6.10^{-3}$ and $SNR = 10$. The squares in the center of figures (a) and (b) represent the sensor boundaries.

6. CONCLUSION

Digital hologram reconstruction is an inversion problem. Classical approaches based on hologram diffraction are not satisfactory because of the artefacts that corrupt the 3-D reconstructions. Hologram reconstruction should not be considered as an inverse wave propagation problem, but rather as a 3-D transmittance reconstruction problem. We described two families of approaches for hologram inversion. Depending on the application, one may either choose to detect iteratively the objects, for simple parametric shapes (typically, spherical objects), or globally reconstruct a 3-D distribution, with a regularizing prior, for more complex and general objects.

Digital camera sensors have a (very) limited size compared to holographic plates. It is therefore crucial in the modeling to consider the finite support of the hologram to prevent from strong border effects in the reconstructions. With a correct handling of the hologram support, accurate reconstructions are possible and can even be extended beyond the field of view. It is also possible to model dead or satured pixels, and signal-dependant noise.

Most existing 3-D reconstruction techniques consider a linear hologram formation model and reconstruct real-valued distributions. Sotthivirat and Fessler[35] considered in their pioneering work the inversion of a non-linear model, with a Poisson noise model, and performed the reconstruction of the complex transmittance distribution. Further works should be done in that direction to evaluate the improvement that non-linear models bring in practice compared to easier to handle linear models.

Inverse problems approaches offer the possibility of optimal hologram processing, which is essential to applications in metrology and to high resolution imaging. The computation of the Cramér Rao lower bounds provide an estimate of the achievable resolution in the reconstructions. The use of non-linear iterative reconstruction techniques however makes it difficult to characterize the actual resolution (since it depends on the object itself). Resolution bounds could yet provide a general methodology to choose the parameters of the optical setup.

REFERENCES

[1] Marquet, P., Rappaz, B., Magistretti, P. J., Cuche, E., Emery, Y., Colomb, T., and Depeursinge, C., "Digital holographic microscopy: a noninvasive contrast imaging technique allowing quantitative visualization of living cells with subwavelength axial accuracy," Optics letters **30**(5), 468–470 (2005).

[2] Poon, T. C., Yatagai, T., and Juptner, W., "Digital holography - coherent optics of the 21st century: introduction," Applied Optics **45**(5), 821 (2006).

[3] Dubois, F., Yourassowsky, C., Monnom, O., and Legros, J., "Digital holographic microscopy for the three-dimensional dynamic analysis of in vitro cancer cell migration," JBO **11**(5), 054032 (2006).

[4] Katz, J. and Sheng, J., "Applications of holography in fluid mechanics and particle dynamics," Annual Review of Fluid Mechanics **42**(1) (2010).

[5] Kreis, T. M., [Handbook of Holographic Interferometry, Optical and Digital Methods], Wiley-VCH, Berlin (2005).

[6] Emery, Y., Cuche, E., Colomb, T., Depeursinge, C., Rappaz, B., Marquet, P., and Magistretti, P., "DHM (Digital holography microscope) for imaging cells," Journal of Physics: Conference Series **61**, 1317–1321 (2007).

[7] Denis, L., Fournel, T., Fournier, C., and Jeulin, D., "Reconstruction of the rose of directions from a digital microhologram of fibres," Journal of Microscopy **225**(3), 283–292 (2007).

[8] Hinsch, K. D. and Herrmann, S. F., "Special issue : Holographic particle image velocimetry," Measurement Science & Technology **15**(4) (2004).

[9] Coupland, J. and Lobera, J., "Special issue : Optical tomography and digital holography," Measurement Science and Technology **19**(7), 070101 (2008).

[10] Gire, J., Denis, L., Fournier, C., Thiébaut, E., Soulez, F., and Ducottet, C., "Digital holography of particles: benefits of the'inverse problem'approach," Measurement Science and Technology **19**, 074005 (2008).

[11] Goodman, J. W., [Introduction to Fourier optics], Roberts & Company Publishers (2005).

[12] Pellat-Finet, P., "Fresnel diffraction and the fractional-order fourier transform," Optics Letters **19**(18), 1388–1390 (1994).

[13] Ozaktas, H. M., Arikan, O., Kutay, M. A., and Bozdagt, G., "Digital computation of the fractional fourier transform," Signal Processing, IEEE Transactions on **44**(9), 2141–2150 (1996).

[14] Lefebvre, C. B., Coetmellec, S., Lebrun, D., and Ozkul, C., "Application of wavelet transform to hologram analysis: three-dimensional location of particles," Optics and Lasers in Engineering **33**(6), 409–421 (2000).

[15] Liebling, M., Blu, T., and Unser, M., "Fresnelets: new multiresolution wavelet bases for digital holography," Image Processing, IEEE Transactions on **12**(1), 29–43 (2003).

[16] Murata, S. and Yasuda, N., "Potential of digital holography in particle measurement," Optics and Laser Technology **32**(7 8), 567–574 (2000).

[17] Malek, M., Allano, D., Coetmellec, S., Ozkul, C., and Lebrun, D., "Digital in-line holography for three-dimensional-two-components particle tracking velocimetry," Measurement Science & Technology **15**(4), 699–705 (2004).

[18] Pan, G. and Meng, H., "Digital holography of particle fields: reconstruction by use of complex amplitude," Applied Optics **42**, 827–833 (2003).

[19] Liebling, M. and Unser, M., "Autofocus for digital fresnel holograms by use of a fresnelet-sparsity criterion," Journal of the Optical Society of America A **21**(12), 2424–2430 (2004).

[20] Dubois, F., Schockaert, C., Callens, N., and Yourassowsky, C., "Focus plane detection criteria in digital holography microscopy by amplitude analysis," Optics Express **14**(13), 5895–5908 (2006).

[21] Fournier, C., Ducottet, C., and Fournel, T., "Digital in-line holography: influence of the reconstruction function on the axial profile of a reconstructed particle image," Measurement Science & Technology **15**, 1–8 (2004).

[22] Jacquot, M. and Sandoz, P., "Sampling of two-dimensional images: prevention from spectrum overlap and ghost detection," Optical Engineering **43**, 214 (2004).

[23] Onural, L., "Sampling of the diffraction field," Applied Optics **39**(32), 5929–5935 (2000).

[24] Coupland, J. M., "Holographic particle image velocimetry: signal recovery from under-sampled CCD data," Measurement Science & Technology **15**, 711?717 (2004).

[25] Stern, A. and Javidi, B., "Analysis of practical sampling and reconstruction from fresnel fields," Optical Engineering **43**, 239 (2004).
[26] Dubois, F., Monnom, O., and Yourassowsky, C., "Border processing in digital holography by extension of the digital hologram and reduction of the higher spatial frequencies," Applied Optics **41**(14), 2621–2626 (2002).
[27] Cuche, E., Marquet, P., and Depeursinge, C., "Aperture apodisation using cubic spline interpolation : application in digital holographic microscopy," Optics communications **182**(23), 59–69 (2000).
[28] Soulez, F., Denis, L., Fournier, C., Thiébaut, E., and Goepfert, C., "Inverse problem approach for particle digital holography: accurate location based on local optimisation," Journal of the Optical Society of America A **24**(4), 1164—1171 (2007).
[29] Soulez, F., Denis, L., Thiébaut, E., Fournier, C., and Goepfert, C., "Inverse problem approach in particle digital holography: out-of-field particle detection made possible," Journal of the Optical Society of America A **24**(12), 3708–3716 (2007).
[30] Denis, L., Lorenz, D., and Trede, D., "Greedy solution of ill-posed problems: error bounds and exact inversion," Inverse Problems **25**, 115017 (2009).
[31] Mallat, S. G. and Zhang, Z., "Matching pursuits with time-frequency dictionaries," Signal Processing, IEEE Transactions on **41**(12), 3397–3415 (1993).
[32] Högbom, J. A., "Aperture synthesis with a non-regular distribution of interferometer baselines," Astronomy and Astrophysics Supplement Series **15**, 417 (1974).
[33] Lee, S. H., Roichman, Y., Yi, G. R., Kim, S. H., Yang, S. M., van Blaaderen, A., van Oostrum, P., and Grier, D. G., "Characterizing and tracking single colloidal particles with video holographic microscopy," Optics Express **15**(26), 18275–18282 (2007).
[34] Cheong, F. C., Krishnatreya, B. J., and Grier, D. G., "Strategies for three-dimensional particle tracking with holographic video microscopy," Optics Express **18**(13), 13563–13573 (2010).
[35] Sotthivirat, S. and Fessler, J. A., "Penalized likelihood image reconstruction for digital holography," Journal of the Optical Society of America A **21**(5), 737–750 (2004).
[36] Brady, D. J., Choi, K., Marks, D. L., Horisaki, R., and Lim, S., "Compressive holography," Optics Express **17**(15), 13040–13049 (2009).
[37] Marim, M. M., Atlan, M., Angelini, E., and Olivo-Marin, J. C., "Compressed sensing with off-axis frequency-shifting holography," Optics letters **35**(6), 871–873 (2010).
[38] Marim, M., Angelini, E., Olivo-Marin, J. C., and Atlan, M., "Off-axis compressed holographic microscopy in low-light conditions," Optics Letters **36**(1), 79–81 (2011).
[39] Denis, L., Lorenz, D., Thiébaut, E., Fournier, C., and Trede, D., "Inline hologram reconstruction with sparsity constraints.," Optics Letters **34**(22), 3475—3477 (2009).
[40] Needell, D. and Tropp, J. A., "CoSaMP: iterative signal recovery from incomplete and inaccurate samples," Applied and Computational Harmonic Analysis **26**(3), 301–321 (2009).
[41] Jacquot, M., Sandoz, P., and Tribillon, G., "High resolution digital holography," Optics Communications **190**(1 6), 87–94 (2001).
[42] Stern, A. and Javidi, B., "Improved-resolution digital holography using the generalized sampling theorem for locally band-limited fields," Journal of the Optical Society of America A **23**(5), 1227–1235 (2006).
[43] Garcia-Sucerquia, J., Xu, W., Jericho, S. K., Klages, P., Jericho, M. H., and Kreuzer, H. J., "Digital in-line holographic microscopy," Applied optics **45**(5), 836–850 (2006).
[44] Kelly, D. P., Hennelly, B. M., Pandey, N., Naughton, T. J., and Rhodes, W. T., "Resolution limits in practical digital holographic systems," Optical Engineering **48**, 095801–1,095801–13 (2009).
[45] Stern, A. and Javidi, B., "Space-bandwidth conditions for efficient phase-shifting digital holographic microscopy," Journal of the Optical Society of America A **25**(3), 736–741 (2008).
[46] Fournier, C., Denis, L., and Fournel, T., "On the single point resolution of on-axis digital holography," Journal of the Optical Society of America A **27**(8), 1856–1862 (2010).
[47] Kay, S. M., [Fundamentals of statistical signal processing: estimation theory] (1993).
[48] Dekker, A. J. D. and den Bos, A. V., "Resolution: a survey," Journal of the Optical Society of America A **14**(3), 547–557 (1997).