Blind deconvolution for thin-layered confocal imaging

Praveen Pankajakshan,1,* Bo Zhang,2 Laure Blanc-Féraud,1 Zvi Kam,3 Jean-Christophe Olivo-Marin,2 and Josiane Zerubia1

1Ariana Project-team, INRIA/CNRS, 2004 Route des Lucioles, BP 93, 06902 Sophia-Antipolis Cedex, France
2Quantitative Image Analysis Unit, Institut Pasteur, 25-28 rue du Docteur Roux, 75015 Paris, France
3Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot 76100, Israel
*Corresponding author: ppankaja@sophia.inria.fr

Received 9 January 2009; revised 14 May 2009; accepted 15 May 2009; posted 3 June 2009 (Doc. ID 105374); published 27 July 2009

We propose an alternate minimization algorithm for estimating the point-spread function (PSF) of a confocal laser scanning microscope and the specimen fluorescence distribution. A three-dimensional separable Gaussian model is used to restrict the PSF solution space and a constraint on the specimen is used so as to favor the stabilization and convergence of the algorithm. The results obtained from the simulation show that the PSF can be estimated to a high degree of accuracy, and those on real data show better deconvolution as compared to a full theoretical PSF model. © 2009 Optical Society of America

100.1455, 180.1790, 100.3190.

1. Introduction

Most of the fluorescence microscopes that image a uniformly illuminated three-dimensional (3D) object by the optical sectioning technique are affected by some out-of-focus fluorescence contributions. Secondary fluorescence from the sections away from the region of interest often interferes with the contrast and resolution of those features that are in focus. Let us take the case of a single-photon (1-p) fluorescence microscope, such as the wide-field microscope (WFM) or the confocal laser scanning microscope (CLSM) [1]. For the sake of simplicity, if we assume that the detectors are the same, then a WFM could be seen as a CLSM but with a fully open pinhole. The WFM can collect more light even from the deeper sections of a specimen but the data are sometimes rendered useless as there is a significant amount of out-of-focus blur. The maximum intensity in each plane decreases as $z^{-2}$, with $z$ being the axial distance from the source. A completely closed pinhole (diameter <1 Airy units (AU); 1AU $= \frac{1.22 \lambda_{ex}}{NA}$), where NA is numerical aperture, on the other hand, confines the light detected only to the in-focus plane but at the expense of imaging low-contrast, highly noisy (signal dependent noise) images. The intensity from a point source in this case decreases as $z^{-4}$ and the loss of in-focus intensity inhibits imaging of weakly fluorescent specimens. Even with a usable pinhole diameter of 1AU, 30% of the light collected is from the out-of-focus regions. In addition, the microscope is inherently diffraction limited [1,2] and the image of a point source (the point-spread function (PSF)) displays a lateral diffractive ring pattern (expanding with defocus) introduced by the finite-lens aperture.

Let $\mathcal{O}(\Omega) = \{o \in \Omega \subseteq \mathbb{R}^{3} \rightarrow \mathbb{R}\}$ denote all possible observable objects on the discrete spatial domain $\Omega = \{(x, y, z) : 0 \leq x \leq N_x - 1, 0 \leq y \leq N_y - 1, 0 \leq z \leq N_z - 1\}$ and $h : \Omega \rightarrow \mathbb{R}$ the microscope PSF. If we assume that the imaging system is linear and shift invariant, then the interaction between $h$ and $o$ is a "3D convolution": $(h \ast o)(x) = \sum_{x' \in \Omega} h(x-x')o(x')$. From the perspective of computational methods, this...
could be inverted with the knowledge of the scanning system properties and also by information about the object being scanned. It is for this reason that the knowledge of the PSF \( h \) is of fundamental importance. The nature of the PSF for fluorescence microscopes has been studied extensively [3–5]. We will introduce the reader, in Subsection 2.B, to one such theoretical model based on the scalar diffraction theory and to its parametric approximation in Subsection 3.B.

A. Problem Formulation

Restoration by deconvolution could be achieved by using either a nonblind or a blind approach. For the nonblind case, the most common approach is an experimental procedure [6,7] that obtains the PSF by imaging a small fluorescent bead (so as to approximate a point object) positioned in the cover slide. Although such a PSF should have been an ideal choice for a deconvolution algorithm, it suffers from low contrast (can be recorded only at finite defocus ranges) and is contaminated by noise. Ways to suppress the noise would be to either acquire several bead data sets and then average them [8,9] or reconstruct them using Zernike polynomial moments [10]. This approach is, however, handicapped by alignment problems and also the whole process could take a long time. The alternative would be to use an analytical model of the PSF [11,12] that takes into account the acquisition system’s physical information as parameters. This information, however, might not be available or might change during the course of the experiment (for example, due to heating of live samples).

We hence arrive at the blind deconvolution approach of estimating the specimen and the unknown PSF parameters using a single observation of the specimen volume. The problem of blind deconvolution is thus reduced to answering the following question: “How does one estimate the original object and the PSF, given only a single observation?”

If we forget the effect of noise and consider the observation model \( (h \ast o) \) in the Fourier space as \( \mathcal{F}(i) = \mathcal{F}(h) \cdot \mathcal{F}(o) \), several solutions for \( o \) and \( h \) answer this problem. For example, if \( (h, o) \) is a solution, then the trivial case is that \( h \) is a Dirac function and \( o = i \) or vice versa. If \( h \) is not irreducible, there exist \( h_1 \) and \( h_2 \) such that \( h = h_1 \ast h_2 \), and the couples \( (h_1 \ast h_2, o) \) and \( (h_1, h_2 \ast o) \) are also solutions. Another ambiguity is in the scaling factor. If \( (h, o) \) is a solution, then \( (\tau h, \tau o) \) \( \forall \tau > 0 \) are solutions, too. This last ambiguity can be waived, for example, by imposing a forced normalization on \( h \). Thus, broadly speaking, a way of reducing the space of possible solutions and to regularize the problem is to introduce constraints on \( h \) and \( o \). If the problem of deconvolution is ill-posed, that of blind deconvolution is underdetermined as the number of unknowns to be estimated is increased without any increase in the input observation data.

Many methods use an iterative approach to estimate the PSF and the object with no prior information on the object [13,14]. Markham and Conchello [15] worked on a parametric form for the PSF and developed an estimation method utilizing this model. The difficulty in using this model for our application is that the number of free parameters to estimate is large and the algorithm is computationally expensive. Hom et al. [16] proposed a myopic deconvolution algorithm that alternates between iteration to deconvolve the object and estimate the PSF. In order to myopically reconstruct the PSF, they introduce a constraint on the optical transfer function (OTF) (the OTF and the PSF are Fourier Transform pairs).

This paper is organized in the following manner. We first discuss the nature of the noise, and then its mathematical modeling and handling in Subsection 2.A. The PSF modeling is introduced in Subsection 2.B. Section 3 is dedicated to the proposed joint restoration and estimation of the imaged object and the microscope PSF using a Bayesian framework. Direct restoration from the observation data is very difficult and, hence, it is necessary to define an underlying model for both the object and the PSF. An alternate minimization (AM) algorithm is then proposed to solve this particular problem. This AM algorithm is then tested on images of degraded phantom objects and real data; the results obtained are presented in Section 4. We then conclude in Section 5 with a discussion and proposed future work. The scope of this paper is restricted to restoring images from a CLSM given the spatial invariance nature of the diffraction-limited PSF.

2. Sources of Distortion and Their Modeling

A. Poissonian Assumption

In digital microscopy, the source of noise is either the signal itself (so-called “photon shot noise”), or the digital imaging system. By tracking the photon-to-electron conversion at the detector, we can observe that the signal and the dependent noise follows an underlying distribution which is Poissonian [17]. Conversely, the imaging noise isolated in the absence of any fluorescence source follows a Gaussian distribution [18,19]. The interested reader may refer to [1,20] for more details on this subject.

In this paper, we have assumed that there is no readout or dark noise as the photomultiplier tube is operating in the photon-counting mode. When the imaging system has been \textit{a priori} calibrated, there is almost negligible offset in the detector and the illumination is uniform. Thus, if \( \{i(x) : x \in \Omega\} \) (assumed to be bounded and positive) denotes the observed intensity of the volume, for the Poissonian assumption, the observation model can be expressed as

\[
\gamma(x) = \mathcal{P}(\gamma([h \ast o](x) + b(x))), x \in \Omega.
\]  

where \( \mathcal{P}(\cdot) \) denotes a voxelwise noise function modeled as a Poissonian process. \( b : \Omega \rightarrow \mathbb{R} \) is a uniformly distributed intensity that models the low-frequency background signal caused by scattered photons and
autofluorescence from the sample. $1/\gamma$ is known as the photon conversion factor, and $\gamma(i(x))$ is the observed photon at the detector.

B. Theoretical Diffraction-Limited Point-Spread Function Model

Among the enormous literature available on PSF modeling, we highlight the work of Stokseth [11], who obtained the OTF for an aberration-free optical system especially for large defocus. This model was used to study the PSF's under different microscope settings and also in validating the algorithm.

If we consider a converging spherical wave in the object space from the objective lens, the near-focus amplitude distribution $h_A$ can be written in terms of the amplitude OTF, OTF$_A$, as $h_A(x) = \int\int$ OTF$_A(k) \exp(jk \cdot x)dk$, where $j^2 = -1$, and $x$ and $k$ are the 3D coordinates in the image and the Fourier space, respectively. By making the axial Fourier space coordinate $k_x$, as a function of lateral coordinates, $k_x = (k_x^2 - (k_y^2 + k_z^2))^{1/2}$, the 3D Fourier transform is reduced to $h_A(x,y,z) := \int\int P(k_x,k_y,k_z) \exp(j(k_x x + k_y y + k_z z))dk_xdk_ydk_z$, where $k = 2\pi\mu/\lambda$ is the wavenumber of an illumination wave with a wavelength $\lambda$ in vacuum and in a medium of refractive index $\mu$, and $P(\cdot,\cdot,\cdot)$ describes the overall complex field distribution in the pupil of an nonaberrated objective lens [2,11].

For an aberration-free microscope, the pupil function can be written as $P(k_x,k_y,z) = \{ A(\phi) \exp(jk\psi), \text{ if } k_x^2 + k_y^2 < k \sin\phi_{\text{max}}, 0, \text{ otherwise} \}$

(2)

where $\psi$ is the optical distance between the wavefront emerging from the exit pupil and the reference sphere measured along the extreme ray, $\phi = \sin^{-1}(k_x^2 + k_y^2)^{1/2}/k$, and $\phi_{\text{max}}$ is the maximum semi-aperture angle of the objective. The intensity projected from an isotropically illuminating point source, such as a fluorophore, on a (flat) pupil plane is bound to be energy conserving. Therefore, the amplitude $A(\phi)$ in the pupil plane for detection should vary as $(\cos \phi)^{-1/2}$ and the energy as $(\cos \phi)^{-1}$ [21]. Conversely, for the illumination case, $A(\phi)$ varies as $(\cos \phi)^{1/2}$. Also, for small defocus, $\psi$ in Eq. (2) could be approximated as $\psi = z(1 - \cos \phi)$ [11]. To derive the intensity distribution of a point source in the image space of a CLSM, we make use of the Helmholtz reciprocity theorem. Since, in induced fluorescence, the excitation ($\lambda_{\text{ex}}$) and the emission wavelengths ($\lambda_{\text{em}}$) are different, the confocal PSF can be written as [22]

$h(x) = C|h_A(x;\lambda_{\text{ex}})|^2 \int\int |h_A(x-x_1,y-y_1,z;\lambda_{\text{em}})|^2dx_1dy_1$,

(3)

where C is a scaling factor and $D$ is the backprojected diameter of the circular pinhole. This theoretical model of the PSF does not take into account aberrations and assumes that diffraction effect predominates the aberrations. However, this scalar model could be extended for other aberrations by modifying the pupil function expression in Eq. (2) to also include the additional phase term due to aberrations [20].

3. Bayesian Framework for the Alternate Minimization Blind Deconvolution Algorithm

In this section we will use the Bayesian framework to describe the method for the blind deconvolution.

A. Deconvolution

Since the advent of the nearest-neighbor deconvolution algorithm [3], there have been numerous techniques proposed [23–26] for image restoration applied to microscopy. These assume, however, that the noise is Gaussian and are valid only for images with high SNR. Statistical methods [27,28], on the other hand, are extremely effective when the noise in the acquired 3D image is fairly strong. We propose here one such nonlinear iterative algorithm that, although slightly computationally expensive (in comparison to linear methods), can better restore the lost higher frequencies.

If we accept the Poissonian model approximation of Eq. (1), then the image $i$ can be interpreted as the realization of independent Poisson processes at each voxel. Hence the likelihood can be written as

$Pr(i|h, \theta) = \prod_{x \in i} \frac{[h \ast o](x)^i(x)e^{-[h \ast o](x)}}{i(x)!}$

(4)

where the mean of the Poisson process is given by $[h \ast o](x)$. In all the derivations used henceforth, the background term has been excluded but the algorithm can be modified by changing the above mean to $[h \ast o + b](x)$. The background fluorescence can be determined from the smoothed histogram of a single “specimen-independent” slice, and it is subsequently added to the mean at every iteration of the maximum likelihood (ML) algorithm Eq. (4) for $o$ [20]. As iterative ML methods do not ensure any smoothness constraints, if unchecked, they evolve to a solution that displays many artifacts from noise amplification (for examples see [29]). There are many remedies, such as terminating the iteration (manually or by using a statistical criterion) before the deterioration begins or prefiltering the observation data. One might argue that, by applying a low-pass filter as a preprocessing step before deconvolution (as in [30]), the results are improved in comparison to the deconvolved images with no prefiltering. The deconvolution algorithm applied after denoising is less influenced by the prior term of the object [21]. However, such prefiltering operations might influence the blind deconvolution algorithm as it is not clear how the resulting filtered data is eventually mapped to the original object. The number of iterations for eventual convergence of the deconvolution algorithm also increases and
the final result need not be optimum. Such interventions are thus a post hoc method of regularizing the ill-posed problem as it is a way of bringing some knowledge about the solution $o$. The maximum a posteriori (MAP) algorithm proposed in this paper uses the prior model on the specimen and the PSF but within the Bayesian framework. We are hence able to simultaneously denoise and deconvolve the observation data without making any modifications whatsoever.

By using the Bayes theorem and assuming that $o$ and $h$ are independent, the posterior joint probability is

$$
Pr(o, h|\hat{i}) = \frac{Pr(\hat{i}|o, h) Pr(o) Pr(h)}{Pr(\hat{i})},
$$

(5)

where $Pr(o)$ is the global prior probability on the object and $Pr(h)$ is the global prior on the PSF. The nature of the prior terms and their expressions are discussed in Subsections 3.A.1 and 3.B. The estimates for $o$ and $h$ can be obtained by simultaneously maximizing the joint probability as

$$
(\hat{o}, \hat{h}) = \arg \max_{(o, h)} \{Pr(o, h|\hat{i})\}
= \arg \min_{(o, h)} \{-\log[Pr(o, h|\hat{i})]\}. 
$$

(6)

As $Pr(\hat{i})$ does not depend on $o$ or $h$, it shall hereafter be excluded from all the estimation procedures that involve either $o$ or $h$. The minimization of the cologarithm of $Pr(o, h|\hat{i})$ in Eq. (6) can be rewritten as the minimization of the following energy functional:

$$
J(o, h|\hat{i}) \equiv J_{\text{obs}}(\hat{i}|o, h) + (\lambda_o J_{\text{reg}, o}(o) + \lambda_h J_{\text{reg}, h}(h)).
$$

(7)

$J_{\text{obs}} : \Omega \rightarrow \mathbb{R}$ is a measure of fidelity to the data and it corresponds to the term $Pr(\hat{i}|o, h)$, which is given from the noise distribution. It has the role of pulling the solution toward the observed data, while $J_{\text{reg}, o} : \Omega \rightarrow \mathbb{R}$ and $J_{\text{reg}, h} : \Omega \rightarrow \mathbb{R}$ are the prior terms on the object and the PSF, which ensure smoothness of the solutions. $\lambda_o$ and $\lambda_h$ are positive parameters that measure the trade-off between goodness of fit and the regularity of the solutions. For the Bayesian interpretation of regularization problems, we refer the reader to the paper by Demoment [32].

Practically, simultaneous estimation of $o$ and $h$ from Eq. (6) is a difficult task. A way to overcome this difficulty is to alternatively maximize the posterior first with respect to $o$ while assuming that the PSF $h$ is known and fixed, and then update the PSF using the previous object estimate. This joint optimization algorithm is summarized as

$$
\hat{o}^{(n+1)} = \arg \max_{o} \{Pr(\hat{i}|o, \hat{h}^{(n)}) Pr(o)\},
$$

$$
\hat{h}^{(n+1)} = \arg \max_{h} \{Pr(\hat{i}|\hat{o}^{(n+1)}, h) Pr(h)\}. 
$$

(8)

The implementation strategy of this blind deconvolution schema has been shown in Algorithm 1 in Subsection 4.A and the discussion follows in the subsequent sections.

1. A Priori Object Models

The ensemble model of an object class refers to any probability distribution $Pr(o)$ on the object space $O$ of the following form:

$$
Pr(o) = Z_{\lambda_o}^{-1} e^{-\lambda_o E(o)},
$$

(9)

where $E(o)$ is a generalized energy and $1/\lambda_o$ (with $\lambda_o > 0$) is the Gibbs parameter for the prior term. We associate with each site $(x, y, z) \in \Omega$ of the object a unique neighborhood $\eta_{xyz} \subseteq \Omega \setminus \{(x, y, z) \in \Omega\}$ as the neighborhood system. If we assume that the random field $(O = o)$ on a domain $\Omega$ is Markovian with respect to the neighborhood system $\eta$, then $Pr(\eta_{xyz}|o_{O \setminus \{(x, y, z) \in \Omega\}}) = Pr(\eta_{xyz}|o_{\eta_{xyz}})$. $o$ is a Markov random field (MRF) on $(\Omega, \eta)$, if $o$ denotes a Gibbs ensemble on $\Omega$ and the energy is a superposition of potentials associated to the cliques (a set of connected pixels). Hence, $E(o) = \sum_{C \in \eta} V_C(o)$.

We use in this paper the following first-order, homogeneous, isotropic MRF, over a six-member neighborhood $\eta_k \in \eta$ (see Fig. 1) of the site $x \in \Omega$:

$$
Pr(O = o(x)) = Z_{\lambda_o}^{-1} e^{-\lambda_o \sum_{x} |\nabla o(x)|},
$$

(10)

where $|\nabla o(x)|$ is the potential function and $\lambda_o$ is the regularization parameter described above. The estimation of this parameter is dealt with in Subsection 4.A.

From a mathematical perspective, $|\nabla o(x)|$ is not differentiable in zero. An approach to circumvent this problem is to regularize it, and instead to consider the (isotropic) discrete definition as

$$
|\nabla o(x,y,z)|_c = ((o(x+1,y,z) - o(x,y,z))^2 + (o(x,y+1,z) - o(x,y,z))^2 + (o(x,y,z+1) - o(x,y,z))^2 + \varepsilon^2)^{1/2},
$$

(11)

where $\varepsilon$ is an arbitrarily small value ($<10^{-3}$). For the partition function $Z_{\lambda_o} = \sum_{o \in C(\Omega)} \exp(-\lambda_o \sum_x |\nabla o(x)|_c)$ to be finite, we restrict the possible values of $o(x)$ so that the numerical gradient of $\nabla o(x)$ is also bounded. When this model is used as a prior for the object, we have the following smoothed regularization functional:
Richardson–Lucy algorithm with total variation regularization

The Euler–Lagrange equation for minimizing \( \mathcal{J}(o, \hat{h}|i) \) in Eq. (14) with respect to \( o \) is

\[
1 - \hat{h}(-x) * \left( \frac{i(x)}{(\hat{o})^n * \hat{h})(\mathbf{x})} - \lambda \text{div} \left( \frac{\nabla o(x)}{|\nabla o(x)|} \right) \right) = 0, \quad (15)
\]

where \( \hat{h}(-x) \) is the Hermitian adjoint operation on \( \hat{h}(x) \) and \( \text{div} \) stands for the divergence (see [28] for details). Inspired by the Richardson–Lucy (RL) algorithm [37,38], Eq. (15) can be solved for the object \( o \) by the following fixed-point iterative algorithm:

\[
\hat{o}^{(n+1)}(x) = \left[ \frac{i(x)}{(\hat{o}^n) * \hat{h})(x)} \right] * \hat{h}(-x) \cdot \frac{\hat{o}^n(x)}{1 - \lambda \text{div} \left( \frac{\nabla \hat{o}^n(x)}{|\nabla \hat{o}^n(x)|} \right)}, \quad (16)
\]

where \((\cdot)\) denotes the Hadamard multiplication (component wise) and \( n \) is the iteration number for the deconvolution algorithm. Equation (16) is similar to the expectation-maximization (EM) algorithm [39] with an underlying statistical model of the process, and can be used for obtaining the MAP estimate of the object. The term \( \text{div} \left( \nabla \hat{o}^n(x) / |\nabla \hat{o}^n(x)| \right) \) can be numerically implemented with the use of central differences and the \text{mod} scheme [29].

Positivity and flux constraint for the object estimate

The deconvolution algorithm that was described above suffers from an inherent weakness. For large values of \( \lambda \), even when the starting guess \( \hat{o}^n \) (with \( n = 0 \)) is positive, the successive estimates need not necessarily have positive intensities. We know that the true intensity of the object \( o(x) \) is always non-negative. Most algorithms truncate these negative intensities to zero or a small positive value. This, however, is a crude manner of handling the estimated intensities as it can lead to loss of some essential information and sometimes also introduces bias into the calculations.

So how else can the problems associated with negative intensity estimates be handled? Fortunately, the problem is entirely due to poor statistical methodology. The modification that we suggest is to include this knowledge of nonnegative true intensities into the prior term of Eq. (10). The distribution that would express precisely this condition is

\[
\Pr(o(x)) := \begin{cases} Z_{\lambda}^{-1} e^{-\sum x |\nabla o(x)|} & \text{if } o(x) \geq 0, \\ 0 & \text{otherwise}. \end{cases} \quad (17)
\]

For the sake of numerical differentiability, we approximate Eq. (17) using a \text{sigmoid} function as

\[
\eta_{\lambda}(x) := \frac{1}{1 + e^{-\lambda x}}.
\]

As in Eq. (6), by applying the \(-\log\) operator to the \( a \) posteriori above, the cost function \( \mathcal{J}(o, \hat{h}|i) \) to be minimized with respect to \( o \) becomes

\[
\mathcal{J}(o, \hat{h}|i) \equiv \sum_{x \in \Omega} [\hat{h} * o](x) - \sum_{x \in \Omega} i(x) \log[\hat{h} * o](x) + \sum_{x \in \Omega} \log(i(x)!)) + \lambda \sum_{x \in \Omega} |\nabla o(x)| + \log[Z_{\lambda}]. \quad (14)
\]
\[
\Pr[\theta(x)] := Z_{\text{new,}\lambda_0}^{-1} e^{-\lambda_0 \sum_x \mathcal{V}_o(x)} \cdot \left( \frac{1}{1 + \exp(\beta_o(\epsilon - o(x)))} \right),
\]

where \( \epsilon \) is a small value close to zero and \( \beta_o \) is a value that specifies the steepness of the sigmoid curve. Typically, the values of \( \beta_o \) and \( \epsilon \) are chosen to be very large and small, respectively, as precision allows. Their values do not individually affect the algorithm and hence need not be known accurately. The cost function of Eq. (14), the Euler–Lagrange Eq. (15), and the multiplicative algorithm in Eq. (16) are thus modified as follows:

\[
\mathcal{J}(o, \hat{h}; \beta) \equiv \left( \sum_{x \in \Omega} |\hat{h} * o(x) - \sum_{x \in \Omega} i(x) \log[\hat{h} * o(x)] \right) + \lambda_o \sum_{x \in \Omega} \mathcal{V}_o(x) + \log[Z_{\text{new,}\lambda_0}]
\]

\[
- \log \left( \frac{1}{1 + \exp(\beta_o(\epsilon - o(x)))} \right),
\]

\[
1 - \frac{\hat{h}(-x) \ast \left( \frac{i(x)}{\hat{h}(-x) \ast \hat{h}(-x)} \right) - \beta_o \exp(\beta_o(\epsilon - o(x)))}{1 + \exp(\beta_o(\epsilon - o(x)))}
\]

\[
= 0.
\]

\[
\partial^{(n+1)}(x) = \left[ \frac{i(x)}{\partial^{(n)}(-x) \ast \hat{h}(-x)} \right] \cdot \frac{\partial^{(n)}(x)}{1 - \lambda_o \exp(\beta_o(\epsilon - o(x)))}
\]

\[
= 0.
\]

Intuitively, the cost function in Eq. (19) ensures that the energy for negative intensity pixels (\( o(x) < \epsilon \)) is very high and, hence, is not reachable (or is not a possible solution) during the iteration procedure.

If the PSF is normalized such that \( ||\hat{h}(x)||_1 = 1 \), in the absence of a background signal, it is simple to show that, for each iteration of the RL algorithm (see Eq. (16) with \( \lambda_0 = 0 \)), the following property is true: \( s = \sum_{x \in \Omega} i(x) = \sum_{x \in \Omega} \partial^{(n)}(x) \). This property is known as the flux or global photometry conservation and it guarantees that the total number of counts of the reconstructed object is the same as the total number of observation counts. However, this property is lost with regularization and can be incorporated by modifying the cost function in Eq. (14) to an additive form or by enforcing it in the following manner after every iteration (except if the background signal is nonzero):

\[
\partial^{(n+1)}(x) = (s^{(0)} \ast \partial^{(n+1)}(x))/s^{(n+1)},
\]

where

\[
s^{(n+1)} = \sum_{x \in \Omega} \partial^{(n+1)}(x),
\]

\[
s^{(0)} = \sum_{x \in \Omega} \partial^{(0)}(x) = \sum_{x \in \Omega} \partial(x).
\]

B. Parametrization of the Point-Spread Function

When \( \lambda_0 = 0 \) in Eq. (14), theoretically speaking, the estimation method on the object and PSF should be the same because \( h \) and \( o \) play a symmetric role. When no constraint is imposed on the PSF, the solution is not always unique. Some reason that a regularization model on the PSF (\( J_{\text{reg}, h} \)) could also be argued along the same lines as the constraints introduced earlier for \( o \) [14,40]. First, a TV [35] kind of regularization cannot model the continuity and regularity in the PSF. A \( \lambda_1 \) kind of norm is suitable only for PSFs that have edges, such as motion blur [41]. Second, in such cases the recovered PSF will be very much dependent on the object/specimen [42]. Separation of the PSF and the object in this case becomes difficult as they have the same or similar solution spaces. Finally, the regularization parameter \( \lambda_1 \) for such a model is highly dependent on the amount of defocus, and varies drastically from one image sample to another. It is for these reasons that we are proposing to intrinsically regularize the PSF through a parametric model.

Because of the invariance property of ML estimate, we can say \( \hat{h}(x) = h(x; \theta_{\text{ML}}) \) is the ML estimate of the PSF. \( \theta \in \Theta \subset \mathbb{R}^+ \) is the set of parameters that defines the PSF. In a more general manner, any PSF can be written as the decomposition on a set of basis functions \( \Phi \) as \( h(x) = \sum_{i=1}^{N_\theta} w_i \Phi_i(x) = (w, \Phi(x)) \), \( \forall x \in \Omega \), where \( w \) denotes the corresponding weights and \( N_\theta \) denotes the number of the basis functions. The imperfections in an image-formation system normally act as passive operations on the data, i.e., they neither absorb nor generate energy. Thus, when an object goes out of focus, it is blurred but the volume's total intensity remains constant. Consequently, all energy arising from a specific point in the fluorescent specimen should be preserved and \( ||h(x)||_1 = \sum_{x \in \Omega} |h(x)| = 1 \). From Eq. (3), it is clear that the intensity distribution of a point source will always be positive and so \( h(x) \geq 0, \forall x \in \Omega \). To satisfy the above-defined conditions, and an additional criterion of circular symmetry (i.e., \( h(-x, -y) = h(x, y), \forall x, y \in \mathbb{R}^2 \)), the Gaussian kernel is chosen as the basis (see [43] for the two-dimensional case). This drastically reduces the number of free parameters to estimate and yet retains a reasonable fit to the actual PSF. It was demonstrated by Zhang et al. [5] that, for a CLSM, a 3D separable Gaussian model gives a relative squared error (RSE) of <9% for a pinhole diameter \( D < 3 \) AU and when the PSF peaks are matched (i.e., \( ||h(x)||_1 = 1 \)), where we say \( \text{RSE} := ||\text{PSF} - h||_2^2/||\text{PSF}||_2^2 \).

Thus the diffraction-limited PSF (with restrictions on the pinhole diameter \( D \)) can be approximated as

\[
h(x) = (2\pi)^{-3/2} |\Sigma|^{-1/2} \exp \left( -\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right),
\]
where \( \mu = (\mu_x, \mu_y, \mu_z)^T \) is the mean vector, \( \Sigma = \begin{bmatrix} \sigma_{11} & \ldots & \sigma_{1z} \\ \vdots & \ddots & \vdots \\ \sigma_{z1} & \ldots & \sigma_{zz} \end{bmatrix} \) is the covariance matrix, and \( \mu_l, \sigma_{l \cdot} \in \Theta \).

As a first approximation, for thin-layered specimen imaging with no aberrations, the PSF is spatially invariant and \( \mu = \{0\} \). A mirror symmetry about the central \( xy \) plane results in a diagonal covariance matrix and, hence, its determinant is \( |\Sigma| = \sigma^2 r^2 z^2 \), where \( \sigma_{11} = \sigma_r \) and \( \sigma_{zz} = \sigma_z \) are the lateral and axial spreads, respectively. It can be shown that the parameters that we are interested in estimating \( \theta = \{\sigma_r, \sigma_z\} \), are dependent on the following settings: wavelength \( \lambda_{ex} \) refractive index \( \mu \), and the NA [5].

1. **Point-Spread Function Parameter Estimation on the Complete Data**

The method outlined in Subsection 3.A.2 requires the knowledge of the PSF \( h(x) \) or \( h(x; \theta) \). From Eqs. (4), (6), and (8), and with the invariance property of ML estimation described earlier, minimizing the energy function with respect to the PSF \( (J(o, h|\theta)) \) or the parameters \( (J(\hat{\theta}, h|\theta)) \) are equivalent. Thus

\[
J(\theta, \hat{o}|\theta) = -\sum_{x \in \Omega} (i(x) \log[h(\theta) + \hat{o}](x)) + \sum_{x \in \Omega} [h(\theta) + \hat{o}](x). \tag{23}
\]

If the true object \( o \) is assumed to be known a priori as \( \hat{o} \), then estimation of the true parameters of the PSF is straightforward as the cost function in Eq. (23) is convex in the neighborhood of optimal \( \theta \in \Theta \) (see Fig. 2). The parameters of the PSF can hence be obtained by a gradient-descent (GD) algorithm [44]. Analytically minimizing Eq. (23) with respect to the parameters leads us to the following:

\[
\hat{\theta}_{l}^{(n+1)} = \hat{\theta}_{l}^{(n)} - \alpha^{(n)} \nabla_{\theta_l} J(\hat{o}, \hat{\theta}_{l}^{(n)}|\theta) \quad \theta_l = \{\sigma_r, \sigma_z\}, \tag{24}
\]

where \( \alpha^{(n)} \) and \( \nabla_{\theta_l} J(\hat{o}, \hat{\theta}_{l}^{(n)}|\theta) \) are the step size and the search direction at iteration \( n \). The gradient of the cost function with respect to the parameters can be calculated as

\[
\nabla_{\theta_l} J(\hat{o}, \hat{\theta}_{l}^{(n)}|\theta) = \sum_{x \in \Omega} \frac{\partial}{\partial \theta_l} h(\theta) + \hat{o} \frac{i(x)}{h(\theta) + \hat{o}} - \sum_{x \in \Omega} \frac{i(x)}{h(\theta) + \hat{o}} \frac{\partial}{\partial \theta_l} h(\theta) + \hat{o} \frac{i(x)}{h(\theta) + \hat{o}}, \quad \forall \theta_l > 0 \in \Theta \tag{25}
\]

If we assume that the PSF is axially and radially centered, i.e., \( \mu = 0, \frac{\partial h(\theta)}{\partial \theta} |_{\mu = 0} = (\frac{-2}{\sigma_r} + \frac{r^2}{\sigma_r^3}) h(\theta), \) and \( \frac{\partial h(\theta)}{\partial \theta} |_{\theta = \sigma} = (\frac{-1}{\sigma_r} + \frac{z^2}{\sigma_z^3}) h(\theta) \). The separable nature of the Gaussian distribution reduces the complexity of the algorithm, as the convolution with the 3D Gaussian PSF can be implemented as three successive 1D multiplications in the Fourier domain. Only a single FFT of the object estimate \( \hat{o} \) needs to be performed as an analytical closed form expression for the Fourier transform of the Gaussian and its derivative exists and can hence be numerically calculated. We stop the computation if the difference measure between two successive iterations is smaller than \( \varepsilon \) (in practice 10\(^{-3}\) or 10\(^{-4}\)), and use the last estimate as the best one.

4. Results

In this section, we validate the proposed AM algorithm on some synthetic and real data.

A. Algorithm Analysis

The global procedure alternatively minimizes the cost function in Eq. (14) first with respect to \( o \)
Eq. (16) while keeping the PSF $h$ fixed and then update the PSF in Eq. (24) using the previous object estimate $o$. Since the iterative algorithm requires an initial guess for the true object, we use the mean of the observed image (i.e., every site is assumed to have a uniform intensity and is, hence, equally likely) for the initialization. For the PSF, as there are no constraints on its spread or support, initialization of the parameters to small values cannot guarantee its convergence to the desired size (due to the Dirac trivial solution). To avoid this problem, we choose the initial parameters to be utmost $2\kappa^{-1}$ resolutions and $6\kappa^{-1}$ resolutions ($1\text{Resel} = 0.61\lambda_{\text{ex}}/\text{NA}$, $\kappa = 2.35$) for the lateral case and the axial case, respectively, and descend down to the optimal value. Both $J_{\text{obs}}(i|o,h)$ and $J_{\text{reg},\rho}(o)$ in Eq. (19) are convex, though not in the strict sense. Although the convergence of the algorithm to the optimal solution is theoretically difficult to prove, numerical experiments indicate that the global procedure does converge when the initialization is carried out as described above.

A delicate situation is in the choice of the regularization parameter $\lambda_o$; too small values yield overly oscillatory estimates owing to noise or discontinuities, while too large values yield overly smooth estimates. The selection or estimation of the regularization parameter is thus a critical issue on which there have been several proposed approaches [45]. However, we are looking for a simple technique that could be combined with the AM algorithm and also fits well with the Bayesian framework. The difficulty in performing marginalization with respect to $\lambda_o$ is that the partition function is not easily computed. An approach to circumvent this problem is by approximating the partition function $Z_{\text{new},\lambda_o}$ as $\lambda_o^{-N_xN_yN_z}$ [46]. By assuming a uniform hyperprior on $\lambda_o$ and maximizing Eq. (19) with respect to $\lambda_o$ leads to the optimal $\lambda_o$ at iteration $(n+1)$ as $\lambda_o^{(n+1)} = (N_xN_yN_z)/\sum_{x=1}^{\lambda_o} |\nabla \hat{o}^{(n)}(x)|$.

B. Numerical Experiments

For the numerical experiments in Fig. 3, we have used a 3D simulated test object of dimensions $128 \times 128 \times 64$, with XY and Z pixel sizes of 20 nm and 50 nm, respectively. The observed data was then generated by using an analytical model of the microscope PSF Eq. (3) (with a pinhole diameter of 1 AU), and the noise was modeled as Poisson statistics [see Fig. 3(b); peak signal-to-noise ratio (PSNR), 16.77 dB]. The results of the AD algorithm are illustrated in Figs. 3(c) and 3(d). The stopping threshold $c$ between two successive iterations was fixed as $10^{-4}$. Figure 4 shows the reduction in the cost function with iterations of the GD algorithm and the approach of the estimated lateral spread parameter $\sigma_r$ to the stable value given the estimate of the object. The quality of the restoration can be assessed by comparing with the original synthetic object using the I-divergence or generalized Kullback distance [29]. For the AM algorithm, when the stopping criterion $c$ was reached, the final I-divergence between $o$ and $\hat{o}$ was 1.4334 (as opposed to 5.55 between $o$ and $I$). Figure 5(a) compares the estimated 3D PSF with the analytically modeled [11] PSF and the best 3D Gaussian fit (in the least-squares sense) for the analytical model. The PSFs are shown along one direction of an off-central lateral plane, and a section of the plot can be viewed as an inset. The maximum of the residual error between the estimate and the true PSF is displayed on a logarithmic contrast stretch in Fig. 5(b). Although the Gaussian model does not capture the ringing sidelobes, as is evident from the residue, the RSE was found to be $<0.07\%$.

C. Experiments on Real Data

1. Imaging Setup and Sample Description

The Zeiss LSM 510 confocal microscope is mounted on a motorized inverted stand (Zeiss Axiovert 200 M) and is equipped with an ArKr excitation laser of wavelength 488 nm. The bandpass (BP) filter transmits emitted light within the band 505–550 nm.

The specimen that was chosen for the first experiment is an embryo of the Drosophila melanogaster [see Fig. 6(a)]. It was mounted and tagged with Green Fluorescent Protein (GFP). This preparation is used for studying the sealing of the epithelial

Algorithm 1: Schema for the Proposed Blind Deconvolution Algorithm

1 begin
2 Input: Observed volume $i \in \mathbb{N}^3$.
3 Data: Initial parameters $\hat{\theta}^{(0)}$ (Subsection 4.A), convergence criterion $\epsilon$.
4 Output: Deconvolved volume $\hat{o} \in \mathbb{N}^3$, PSF parameters $\hat{\theta} \in \Theta \subset \mathbb{R}^2$.
5 Initialization: $n = 0$, $\hat{o}^{(0)}(x) = \text{Mean}(i(x))$, $\hat{h}^{(0)}(x) = h(x; \hat{\theta}^{(0)})$ [Eq. (22)].
6 Estimate the background term $\hat{b}$ from the image histogram (Subsection 3.A).
7 while $|\hat{o}^{(n)} - \hat{o}^{(n-1)}|/\hat{o}^{(n)} > \epsilon$ do
8 $\lambda_o$ estimation: $\lambda_o^{(n)} = 1/\text{Mean}(\nabla \hat{o}^{(n)}(x))$.
9 Using the minmod scheme [29], calculate $\text{div}(\nabla \hat{o}^{(n)}(x)/|\nabla \hat{o}^{(n)}(x)|)$.
10 Deconvolution: Calculate $\hat{\theta}^{(n+1)}$ from Eq. (21).
11 Projection Operation: Scale $\hat{o}^{(n+1)}$ for preserving the flux (Subsection 3.A.2).
12 Parameter estimation: Calculate $\hat{\theta}^{(n+1)}$ from Eqs. (24) and (25).
13 Assign: $\hat{h}^{(n+1)}(x) = h(x; \hat{\theta}^{(n+1)})$ and $n = (n + 1)$.
14 end
15 end
sheets (dorsal closure) midway during the embryogenesis. The objective lens is a Plan-Neofluar with 40× magnification having a NA of 1.3 and immersed in oil (Immersol 518 F, Zeiss, refractive index μ = 1.518). The pinhole size was 67 μm. The images (Institute of Signaling, Development Biology and Cancer, Nice UMR6543/CNRS/UNS) were acquired with a XY pixel size of 50 nm and a Z step size of 170 nm, and the size of the volume imaged is 25.59 μm × 25.59 μm × 2.55 μm.

The second set of images [National Institute for Agricultural Research (INRA), Sophia-Antipolis, France] are the root apex of the plant Arabidopsis thaliana immersed in water (see Fig. 7). The dissected roots of the Arabidopsis thaliana plant were directly put on a microscope slide in approximately 100 μl of water and this was then gently covered with a coverslip. This simple set up works very well when the image acquisition recording times are not too long (about 30 min). The microscope specifications are the same as that used for acquiring the first data set, but the objective is a C-Apochromat water immersion lens with 63× magnification, 1.2 NA. The lateral pixel dimensions are 113 nm and the Z step is 438 nm. The pinhole was fixed at 110 μm. This preparation was used to study Nematode infection at the center of the root in the vascular tissue.

2. Deconvolution Results

A rendered subvolume of the observed and restored data for the Drosophila melanogaster is shown in Fig. 6. The deconvolution algorithm was stopped when the difference between subsequent estimates was lower than ε = 0.002. The AM algorithm converged after 40 iterations of the joint RL–TV and GD algorithm. The PSF parameters were initialized to 300 nm and 600 nm for the lateral and the axial case, respectively, and the GD algorithm estimated them to be 257.9 and 477.9 nm [47]. These are larger (by about 16% and 14.5% for the lateral and the axial cases, respectively) than their corresponding theoretically calculated values [5]. These results are also fully in line with an experimental study performed earlier [48] with subresolution beads,
which indicated a large deviation between theoretical aberration-free PSF models and empirically determined PSFs.

Figure 8(a) shows a rendered subvolume (as indicated in Fig. 7) of the observed root apex and the corresponding restored result is shown in Fig. 8(b). It is evident from these results that the microtubules (as identified by their specific binding proteins, microtubules binding domain (MBD)) are more easily discerned in the restoration than in the original data. It was verified from the experiments on synthetic data [47] that the proposed algorithm can not only estimate the actual PSF, but can also provide a much better deconvolution result [49] in comparison to theoretical microscope PSFs (generated using the microscope settings). Validation is very important as, in some situations, artifacts might arise in the restored image. These artifacts would be hard to distinguish from biological structures unless some knowledge about the true image is available. However, the results on real data are difficult to validate unless a higher resolution image of the same sample is available. Hence, we tested our deconvolution algorithm on images of spherical fluorescent shells (see [36]) whose thickness was measured after deconvolution and found to be closer to the true value specified by Molecular Probes.

5. Conclusions and Future Work

In this paper we have proposed and validated an AM algorithm for the joint estimation of the microscope PSF and the specimen source distribution for a CLSM. We choose the RL algorithm for the deconvolution process as it is best suited for the Poisson data, and TV as the regularization model. A separable 3D Gaussian model best describes the diffraction-limited confocal PSF, and is chosen as the a priori model for the PSF. We are able to achieve blind deconvolution by constraining the solution of the object and the PSF to different spaces. The PSF approximation that is given in this paper is currently relevant to imaging thin samples. However, it could also be extended to encompass any PSF that can be decomposed in a similar manner. We have experimented on simulated and real data, and the method gives very good deconvolution results and a PSF estimation close to the true value [29,47]. However, it should be noted that all of the out-of-focus light cannot be rejected and some noticeable haze and axial smearing remains in the images. This could be improved by adding a Gamma prior on the PSF parameters.

This research was partially funded by the P2R Franco–Israeli Collaborative Research Program and the French National Research Agency (ANR) DETECFINE Project. The authors gratefully acknowledge Dr. Caroline Chaux (Université Paris-Est, France) and Prof. Arie Feuer (Technion, Israel) for several interesting discussions. We would also like to thank INRIA for supporting the Ph.D. of the first author through a state-subsidized INRIA doctoral research contract (CORDI-S) fellowship and CNRS for supporting the Ph.D. of the second author. Additionally, our sincere gratitude goes to...
Dr. Stéphane Noselli, Dr. Fanny Serman from the Institute of Signaling, Development Biology and Cancer UMR 6543/CNRS/UNS, and Dr. Gilbert Engler (French National Institute for Agricultural Research (INRA) Sophia-Antipolis, France) for painstakingly preparing the images presented in Subsection 4.C, and for the useful comments on their validation.

References


1 August 2009 / Vol. 48, No. 22 / APPLIED OPTICS 4447


