Parameter-Free Gaussian PSF Model for Extended Depth of Field in Brightfield Microscopy

Xu Zhou, Rafael Molina, *Senior Member, IEEE*, Yi Ma, *Fellow, IEEE*, Tianfu Wang, and Dong Ni

***Abstract*— Due to their limited depth of field, conventional brightfield microscopes cannot image thick specimens entirely in focus. A common way to obtain an all-in-focus image is to acquire a z-stack of images by optically sectioning the specimen and then apply a multi-focus fusion method. Unfortunately, for undersampled image stacks, fusion methods cannot remove the blur in regions where the in-focus position is between two optical sections. In this work, we propose a parameter-free Gaussian PSF model in which the all-in-focus image together with both the depth map and sampling distances in image plane are estimated from the image sequence automatically, without knowledge on the z-stack acquisition. In a maximum a posteriori framework, an iteratively reweighted least squares method is used to estimate the image and an adaptive scaled gradient descent method is utilized to estimate the depth map and sampling distances efficiently. Experiments on synthetic and real data demonstrate that the proposed method outperforms the current state-of-the-art, mitigating fusion artifacts and recovering sharper edges.**

***Index Terms*— Blind deconvolution, point spread function, shape from focus, focal stack.**

I. Introduction

IN microcopy, the limited depth of field of optical systems is an important problem which must be addressed. It is impossible to capture an image in which all the objects are in focus. In other words, only portions of the scene within the system's focal range are in focus, while others exhibit various amount of blur. One common remedy for conventional brightfield microscopy is to acquire a sequence of optical sections of the sample by gradually moving the sample along the optical axis, generating a z-stack of images that contains all the available in-focus information on the specimen.

Manuscript received February 21, 2019; revised August 12, 2019, October 22, 2019, and November 16, 2019; accepted November 19, 2019. Date of publication December 11, 2019; date of current version January 28, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 61571304, in part by the Shenzhen Peacock Plan under Grant KQTD2016053112051497 and Grant KQJSCX20180328095606003, and in part by the Medical Scientific Research Foundation of Guangdong Province, China, under Grant B2018031. The work of R. Molina was supported by the Spanish Ministerio de Economía y Competitividad under Contract DPI2016-77869-C2-2-R. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Xiaochun Cao. *(Corresponding author: Dong Ni.)*

X. Zhou, T. Wang, and D. Ni are with the School of Biomedical Engineering, Shenzhen University, Shenzhen 518060, China (e-mail: zhouxu179@163.com; tfwang@szu.edu.cn; nidong@szu.edu.cn).

R. Molina is with the Departamento de Ciencias de la Computación e Inteligencia Artificial, E. T. S. I. Informática y Telecomunicación, Universidad de Granada, 18071 Granada, Spain (e-mail: rms@decsai.ugr.es).

Y. Ma is with the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, Berkeley, CA 94720-1770 USA (e-mail: yima@eecs.berkeley.edu).

This article has supplementary downloadable material available at http://ieeexplore.ieee.org, provided by the authors.

Digital Object Identifier 10.1109/TIP.2019.2957941

Mathematically, the z-stack formation model can be described in matrix-vector form as [1]–[3]

$$g_l = H_l f + n_l, \quad l \in [1, L], \tag{1}$$

where $g_l \in \mathbb{R}^N$ is the lth observation in the z-stack, $n_l \in \mathbb{R}^N$ is assumed to be zero-mean Gaussian white noise, $f \in \mathbb{R}^M$ represents the latent all-in-focus image, and H_l denotes the $M \times N$ blur matrix corresponding to the lth capture, which is defined by the Point Spread Function (PSF) of the system. In this work, we aim to estimate the latent all-in-focus image f from a given set of blurred images $\{g_1, g_2, \dots, g_L\}$, without knowing the sequence of blurs $\{H_1, H_2, \dots, H_L\}$.

To estimate a sharp image from a z-stack, multi-focus image fusion [4]–[11] or model based deconvolution [12], [13] methods can be used. Most multi-focus image fusion methods obtain the all-in-focus image as a weighted average of the z-stack images, either in the spatial or a transformed domain, see [14] for a recent literature review. The quality of the fused image mainly relies on the focus measure (FM) used. This measure determines the weighting values. Numerous FMs have been proposed in the last decade, see [5] for a comprehensive analysis. Commonly used FMs are image gradient (first or high order derivatives), image variance, wavelet coefficients, discrete cosine transform coefficients and image statistics. FM weights strongly depend on the window size used. The fused image often exhibits some blocking artifacts, since an image block may contain both in focus and defocus areas. Pixel based multi-focus fusion methods, including guided filter fustion (GFF) [6], dense scale-invariant feature transform (DSIFT) [8], dictionary based sparse representation [7], directional-max-gradient flow with depth propagation [9], and convolutional neural networks (CNN) [10], [11] better preserve spatial consistency and mitigate block artifacts. Unfortunately, depending on the depth of field (DOF) of the optical system and the thickness of specimen, multi-focus fusion methods often require tens or even hundreds of partially in focus images to generate an all-in-focus image [15]. If the z-stack is undersampled, multi-focus fusion methods cannot remove the defocus blurs in every image sample due to the averaging operation. This weakness of fusion methods restricts their use in situation where image acquisition cost (time) is limited, e.g., cervical cancer screening.

Model based deconvolution methods take into account the z-stack formation model and iteratively recover both the latent sharp image and the depth map. For an undersampled z-stack, the resulting problem becomes ill-posed, and priors on the latent image and the depth map must be introduced to remove defocus blurs. When the optical constant and sampling distances are both given, Auget *et al.* [12] propose a 2.5-D Gaussian blur model for extended depth of field in brightfield microscopy. Provided that the camera constants have already been calibrated, Rajagopalan and Chaudhuri [3] propose to reconstruct the all-in-focus image and depth map by modeling both the restoration and depth varying parameters as Markov random fields. Lin *et al.* [13] first estimate the initial depth map by utilizing the relative blur [16] and total variation regularization, and then alternatively refine the latent image and depth map using sparse image priors [17]. References [3] and [13] need camera calibration for different camera settings and the image acquisition is not performed by optical sectioning as in brightfield microscopy.

Non-model based deconvolution methods [18], [19], and [20], which rely on blur map estimation, have also been proposed to remove depth varying defocus blur from a single image. Unfortunately, due to the missed information during sampling, these blind deconvolution methods perform worse than multi-focus fusion ones, since fine details that have been significantly blurred cannot be recovered. Assuming that blur kernels are constant within a local 9×9 patch, the recent multi-image deconvolution method with ℓ_1 regularization [21] recovers an all-in-focus super-resolution image from a focal stack. However, since this method is not a model based one, the depth map cannot be recovered and the blur kernels are not correlated along the optical axis.

Apart from the above computational methods, optical techniques also can be used to extend the depth of field. The wavefront coding methods [22]–[24] utilize a cubic phase mask, which is placed at (or conjugate to) the exit pupil of the microscope objective, to make the out-of-focus PSF similar to the in-focus PSF. A sharp all-in-focus image is then reconstructed from the captured image using image deconvolution. Levin *et al.* [17] show that a coded aperture and patch based kernels can be used to improve the quality of all-in-focus image. A different approach [25] uses volumetric optical sampling to acquire an accumulated image. It assumes that the blur is spatially invariant and reconstructs the sharp image by deconvolving the captured image with the accumulated PSF.

In this work, following the depth varying Gaussian PSF model in [26], we propose a parameter free Gaussian PSF model in which the optical constant is removed by including it into the unknown depth map and sampling distances. We show that, without knowing the optical constant and z-stack sampling distances, the latent image, depth map and sampling distance (both in the image plane sense) can be estimated from the given z-stack by solving a regularized least-squares minimization problem. Since the proposed model does not need any information on the optical system, it can be used in a wide range of scenarios. In the proposed framework, an iteratively reweighted least squares method is used to estimate the image and a new adaptively scaled gradient descent algorithm is proposed to estimate the depth map and sampling distances efficiently. Experiments on synthetic and real data demonstrate that the proposed method outperforms the current state-of-the-art, mitigating fusion artifacts and recovering sharper edges.

II. Model Formulation

In this section, we first review the two 2.5-D convolution models presented in [12] and [26], respectively. We then discuss both models and finally formulate our parameter-free Gaussian PSF model.

A. 2.5-D Convolution Model With Gaussian PSF

Without loss of generality, we assume that the size of image g_l is $[\sqrt{N}, \sqrt{N}]$ and the largest PSF kernel size is $[2k+1, 2k+1]$. Following the 2.5-D convolution model in [12], the observed image g_l is a shift-variant convolution between the latent sharp image f and the 3D PSF $h(x, y, \nu)$ of the microscope plus noise, that is,

$$g_l(u, v) = n_l(u, v) + \sum_{x=-k}^{k} \sum_{y=-k}^{k} f(u-x, v-y) h(x, y, d(u-x, v-y) - z_l), \tag{2}$$

where $d(u, v)$ denotes the depth map whose value represents the distance between the surface point at pixel (u, v) and the objective lens; z_l represents the sampling distance at which the surface of specimen will be in focus; and the use of $\sqrt{M} = \sqrt{N} + 2k$ indicates that the undermined boundary conditions have been imposed to mitigate boundary artifacts, see [27] for more about boundary conditions. For a given depth $d(u, v)$ and a sampling distance z, the PSF $h(x, y, \nu)$, where $\nu(u, v, x, y) = d(u-x, v-y) - z$ and (x, y) denotes displacement vector, is assumed to be the Gaussian function

$$h(x, y, \nu) = \frac{1}{2\pi(\eta_0 + \eta_1|\nu|)^2} \exp\left(-\frac{(x^2 + y^2)}{2(\eta_0 + \eta_1|\nu|)^2}\right), \tag{3}$$

where η_0 is a small positive number to avoid division by zero and η_1, which depends on z and d, represents the optical constant that converts world coordinates to image plane.

It is shown in [12] that the 2.5-D shift-variant convolution model in Eq. (2) is valid provided that the specimen is opaque or the specimen is transparent but much thicker than the DOF of the system. A sequence of z-stack images can be captured by varying the distance between the specimen and objective lens. This sequence can be approximately described by Eq. (2).

The image formation model in [26] (see Eq. (14), in page 1136) assumes the following simpler form

$$g_l(u, v) = \sum_{x=-k}^{k} \sum_{y=-k}^{k} f(u-x, v-y) h(x, y, \nu_l(u, v)) + n_l(u, v), \tag{4}$$

where $\nu_l(u, v) = d(u, v) - z_l$ and $h(x, y, \nu_l)$ is the Gaussian function in Eq. (3). In general, the image formation model in Eq. (2) is more precise than the one in Eq. (4), since the local kernel takes into account the neighbouring depth information. However, for a smooth surface, both models are numerically close. In this work, assuming that the surface of specimen is smooth, we adopt the image formation model in Eq. (4) since it is mathematically more tractable.

B. Shortcomings of the Gaussian PSF in Eq. (3)

Using the thin lens model for the microscope in [15] and according to the defocus blur radius equation given in [1], [28], we have

$$\eta_1 = \frac{sF^2}{Ad(z-F)}, \tag{5}$$

where s denotes resolution (measured in pixels per micrometer), F is the focal length, and A is the f-number. Since $d(z-F)$ is almost constant (due to very reduced thickness of specimens) and because the variability in $\{z_l\}$ is limited, in [12], η_1 and $\{z_l\}$ are treated as input constants provided by the user.

However, the value of η_1 is difficult to know, since the depth map d is unknown in practice. Consequently, it is hard to provide an accurate estimate of η_1, and even more difficult when $\{z_l\}$ is unavailable. Unfortunately, a wrong estimate of η_1 also introduces error in the estimation of the depth map. If $\eta_1(d - z_l^*) = \eta_1^*(d^* - z_l^*)$ (assuming no error in the PSF for both models) with $*$ indicating groundtruth, then we have the depth estimate error

$$\|d - d^*\| = \|(\frac{\eta_1^*}{\eta_1} - 1)(d^* - z_l^*)\|, \tag{6}$$

which increases proportionally to the relative error of the optical constant.

In addition, the normalization scalar $2\pi(\eta_0 + \eta_1|z|)^2$ in Eq. (3) is true for infinite kernel support only. As a result, a small approximation error which depends on the assumed kernel size always exists in practice. The Gaussian PSF model in [12] is never normalized and the total intensity of the recovered image is modified. In summary, drawbacks of the Gaussian PSF model in [12] include a) the need to know the optical constant η_1 and the sampling distance $\{z_l\}$ and b) the PSF error caused by the finite kernel size and the inaccurate estimates of optical constant and sampling distances. Let us now introduce our model which solves the above indicated problems.

C. Parameter-Free Gaussian PSF

Since we aim to estimate the latent sharp image, we rescale ν, d, and z by multiplying all of them by $\sqrt{\eta_1}$ and define the discrete form of the 3D parameter-free Gaussian PSF as

$$h(x, y, \nu) = \frac{\exp\left(-\frac{(x^2+y^2)}{2(\eta_0+\nu^2)}\right)}{\sum_{m=-k}^{k}\sum_{n=-k}^{k}\exp\left(-\frac{(m^2+n^2)}{2(\eta_0+\nu^2)}\right)}. \tag{7}$$

Notice that η_0 is just a small positive number to avoid division by zero. The proposed Gaussian PSF in Eq. (7) is parameter free. The unknown depth and sampling distances will be estimated from the z-stack. Furthermore, thanks to the denominator in Eq. (7), all blur kernels are normalized for arbitrary kernel support size.

Now substituting Eq. (7) into Eq. (4), we obtain the z-stack model formulation:

$$g_l(u, \upsilon) = \sum_{x=-k}^{k}\sum_{y=-k}^{k} \frac{f(u-x, \upsilon-y)\exp\left(-\frac{(x^2+y^2)}{2(\eta_0+\nu_l^2(u,\upsilon))}\right)}{\sum_{m=-k}^{k}\sum_{n=-k}^{k}\exp\left(-\frac{(m^2+n^2)}{2(\eta_0+\nu_l^2(u,\upsilon))}\right)} + n_l(u, \upsilon), \tag{8}$$

where ν_l is a vector with components $\nu_l(u, \upsilon) = d(u, \upsilon) - z_l$ and the redundant scalar η_1 has been removed for simplicity. In what follows, the depth map d and sampling distances z_l are understood in image plane.

Let $f_{xy}(u, \upsilon) = f(u-x, \upsilon-y) \in \mathbb{R}^N$ be the shifted subimage of f, the matrix-vector convolution in Eq. (1) can be written as

$$H_l f = \sum_{x=-k}^{k}\sum_{y=-k}^{k} f_{xy} \circ h(x, y, \nu_l), \tag{9}$$

where $\circ$ denotes element-wise multiplication.

The benefits of the proposed PSF model in Eq. (7) include a) no need to know the z-stack acquisition and b) the possibility of finding the best sampling distances and depth map that match the model in Eq. (8) regardless of factors like optical diffraction [29], [30], aberration [15], digitalization [31], nonlinear camera response, downsampling, and image compression.

III. Problem Formulation

Given a z-stack $\{g_1, g_2, \ldots, g_L\}$, our goal is to estimate the latent sharp image f, depth map d and sampling distances $\{z_l\}$ using a probabilistic based formulation. To achieve this goal we introduce the priors on the unknown (latent) variables and combine them with the observation model to obtain the cost function to be optimized when the MAP (maximum a posteriori) approach is used.

Assuming an i.i.d. Gaussian noise with variance σ^2, we write the observation model as

$$P(\{g_l\}|f, d, \{z_l\}) \propto \exp\Big(-\sum_{l=1}^{L}\frac{\|H_l f - g_l\|_2^2}{2\sigma^2}\Big). \tag{10}$$

Unfortunately, the maximization of the above likelihood is an ill-posed problem, particularly when the z-stack is undersampled. Priors on the latent image and depth map are indispensable. Levin *et al.* [17] have shown that natural images have a heavy-tail distribution in the gradient domain. We adopt the following sparse prior on the latent image

$$P(f) \propto \exp\big(-\frac{\lambda_f}{\sigma^2}\sum_{\gamma=1}^{5}\|D_\gamma f\|_p\big), \tag{11}$$

where D_γ is the partial convolution operator formed by the derivative filter ∇_γ ($\nabla_1 = [1, -1], \nabla_2 = \nabla_1^T, \nabla_3 = [-1, 2, -1], \nabla_4 = \nabla_3^T, \nabla_5 = [1, -1; -1, 1]$) and p is a scalar in the range $[0.6, 0.8]$ which enforces natural image sparsity in the gradient domain, see [17] for details. See [32] and [33] for additional information on sparsity promoting priors.

Since we are interested in cell images, we impose a smooth prior on the depth map, which is defined as

$$P(d) \propto \exp\left(-\frac{\lambda_d}{\sigma^2}\sum_{\gamma=1}^{2}\|D_\gamma d\|_2^2\right). \tag{12}$$

The depth prior plays an important role in depth recovery and should be chosen properly according to the shape of specimen. For example, if the shape of specimen is piecewise smooth, sparse priors including TV [34], ℓ_p [17] and nonlocal means [26] are good candidates.

Finally, assuming a uniform improper prior for the sampling distances, namely $P(\{z_l\}) \propto const$, we have

$$\begin{aligned} P(f, d, \{z_l\}|\{g_l\}) &\propto P(\{g_l\}|f, d, \{z_l\})P(f)P(d)P(\{z_l\}) \\ &\propto \exp(-\sum_{l=1}^{L}\frac{1}{2\sigma^2}\|H_l f - g_l\|_2^2 \\ &\quad -\frac{\lambda_f}{\sigma^2}\sum_{\gamma=1}^{5}\|D_\gamma f\|_p - \frac{\lambda_d}{\sigma^2}\sum_{\gamma=1}^{2}\|D_\gamma d\|_2^2). \end{aligned} \tag{13}$$

Now that we have the joint distribution, we proceed to calculate the mode of the posterior distribution $P(f, d, \{z_l\}|\{g_l\})$, the so called MAP estimate, which is obtained by maximizing the logarithm of the posterior or equivalently by minimizing the following cost function

$$\begin{aligned} &J(f, d, \{z_l\}) \\ &= \sum_{l=1}^{L}\frac{1}{2}\|H_l f - g_l\|_2^2 + \lambda_f\sum_{\gamma=1}^{5}\|D_\gamma f\|_p + \lambda_d\sum_{\gamma=1}^{2}\|D_\gamma d\|_2^2. \end{aligned} \tag{14}$$

The cost function $J(f, d, \{z_l\})$ consists of three terms: the data fidelity term $J_f = \sum_{l=1}^{L}\frac{1}{2}\|H_l f - g_l\|_2^2$, the image regularization term $R_f = \lambda_f\sum_\gamma\|D_\gamma f\|_p$ and the depth map regularization term $R_d = \lambda_d\sum_{\gamma=1}^{2}\|D_\gamma d\|_2^2$. The data fidelity term takes the dependence on the observation data into account and the regularization terms enforce the desired smoothness. The regularization weights λ_f and λ_d represent the trade-off between data fidelity, image smoothness and depth smoothness. Our goal now becomes the minimization of the cost function.

IV. Optimization

The cost function $J(f, d, \{z_l\})$ is minimized by an iterative alternating optimization scheme: a) fixing the latent image, the depth map and sampling distances are estimated; b) fixing the depth map and sampling distances to estimate the latent image.

A. Joint Depth and Sampling Distance Estimation

Fixing f in $J(f, d, \{z_l\})$, we use gradient descent to minimize the nonconvex cost

$$J_{dz}(d, \{z_l\}) = \sum_{l=1}^{L}\frac{1}{2}\|H_l f - g_l\|_2^2 + \lambda_d\sum_{\gamma=1}^{2}\|D_\gamma d\|_2^2. \tag{15}$$

The gradients $\nabla_d J_{dz}$ and $\nabla_{z_l} J_{dz}$ are computed using the chain rule. First of all, we have

$$\nabla_d J_{dz} = \sum_{x,y,l}\frac{\partial h(x, y, \nu_l)}{\partial d}\circ\frac{\partial J_{dz}}{\partial h(x, y, \nu_l)} + 2\lambda_d\sum_{\gamma=1}^{2}D_\gamma^T D_\gamma d, \tag{16}$$

where $\partial J_{dz}/\partial h(x, y, \nu_l)$ and $\partial h(x, y, \nu_l(u, v))/\partial d(u, v)$ are respectively given by

$$\frac{\partial J_{dz}}{\partial h(x, y, \nu_l)} = f_{xy}\circ(H_l f - g_l) \tag{17}$$

and

$$\begin{aligned} \frac{\partial h(x, y, \nu_l(u, v))}{\partial d(u, v)} &= h(x, y, \nu_l(u, v))\frac{\nu_l}{(\eta_0 + \nu_l{}^2(u, v))^2} \\ &\quad [(x^2 + y^2) - \sum_{m=-k}^{k}\sum_{n=-k}^{k}h(m, n, \nu_l(u, v))(m^2 + n^2)]. \end{aligned} \tag{18}$$

Similarly, we have

$$\begin{aligned} \nabla_{z_l} J_{dz} &= \sum_{x,y,u,v}\frac{\partial h(x, y, \nu_l(u, v))}{\partial z_l}\frac{\partial J_{dz}}{\partial h(x, y, \nu_l(u, v))} \\ &= -\sum_{x,y,u,v}\frac{\partial h(x, y, \nu_l(u, v))}{\partial d(u, v)}\frac{\partial J_{dz}}{\partial h(x, y, \nu_l(u, v))}. \end{aligned} \tag{19}$$

With the gradients in Eqs. (16) and (19), we use the backtracking line search algorithm [35] to minimize the cost J_{dz}. Notice that J_{dz} is expected to change faster on z than on d, to compensate for the differences in variation, we have experimentally observed that a better convergence is obtained when we scale the gradient $\nabla_d J_{dz}$ by $\max(\max_l|\nabla_{z_l} J_{dz}|/\max_{u,v}|\nabla_d J_{dz}|, 1)$.

The Adaptively Scaled Gradient Descent Algorithm (ASGDA) is presented in Alg. 1. It should be noted that z_1 is fixed to 0 to eliminate the ambiguity caused by multiple solutions. Without fixing z_1, there are infinite pairs of optimal solutions $\{d^*+\theta, \{z_l^*+\theta\}\}$ where θ is an arbitrary real number. Numerical results demonstrate that the proposed ASGDA finds an approximate (local or saddle) optimal solution efficiently.

B. Image Estimation

Fixing d and $\{z_l\}$ in $J(f, d, \{z_l\})$, we now minimize the function

$$J_f(f) = \sum_{l=1}^{L}\frac{1}{2}\|H_l f - g_l\|_2^2 + \lambda_f\sum_{\gamma=1}^{5}\|D_\gamma f\|_p, \tag{20}$$

whose stationary points can be approximately found by iteratively reweighted least squares (IRLS). Using IRLS, at iteration t $(t = 0, 1, 2, ..)$ we update the weights using

$$W_\gamma^t = min(p|D_\gamma f^t|^{p-2}, p\epsilon^{p-2}) \tag{21}$$

and then find f^{t+1} by solving the linear system

$$(\sum_{l=1}^{L}H_l^T H_l + \lambda_f\sum_{\gamma=1}^{5}D_\gamma^T diag(W_\gamma^t)D_\gamma)f^{t+1} = \sum_{l=1}^{L}H_l^T g_l, \tag{22}$$

Algorithm 1 Adaptive Scaled Gradient Descent Algorithm

Require: $\{g_l\}$, f, d, $\{z_l\}$, λ_d, tol.

$n = 0$, $z_1 = 0$, $\tau_2 = 0.5$, $lr = 1$.
while $\max_{u,v} |\nabla_d J_{dz}| \geq tol$ **do**
  $scale = \max(\max_{l\geq 2} |\nabla_{z_l} J_{dz}| / \max_{u,v} |\nabla_d J_{dz}|, 1)$
  $\hat{d} = d - lr * scale * \nabla_d J_{dz}$
  $\hat{z}_l = z_l - lr * \nabla_{z_l} J_{dz}$, $(2 \leq l \leq L)$
  while $J_{dz}(\hat{d}, \{\hat{z_l}\}) \geq J_{dz}(d, \{z_l\})$ **do**
    $lr = \tau_2 * lr$
    $\hat{d} = d - lr * scale * \nabla_d J_{dz}$
    $\hat{z}_l = z_l - lr * \nabla_{z_l} J_{dz}$, $(2 \leq l \leq L)$
  end while
  $lr = lr * 1.2$
  $d = \hat{d}$, $z_l = \hat{z}_l$,
end while
return d, $\{z_l\}$

where ϵ is a small positive real number and the linear system can be solved efficiently by the conjugate gradient (CG) method, see [35]. We refer interested readers to [36] for the derivations of Eqs. (21) and (22) and the proof of convergence. Now that we have the optimization procedure to find the latent image, the depth map, and the sampling distances, let us see how we initialize them.

C. Initializations

The cost function $J(f, d, \{z_l\})$ is nonconvex and the proposed optimization method may converge to poor local minima. A good initialization not only reduces the running time but also helps avoid poor local minima.

1) Image Initialization: Numerous multi-focus image fusion methods, including the recent GFF [6], DSIFT [8] and CNN based [10] method, just to name a few, can be used for latent image initialization. For simplicity and speed, and based on the fact that the total variation of a patch decreases as defocus increases [37], we adopt as focus measure, the Patch Total Variation (PTV), which is similar to Spatial Frequency Measure [5], and it is defined as

$$PTV_l = AP_{re}\sqrt{(D_1 g_l)^2 + (D_2 g_l)^2}, \quad l \in \{1, \ldots, L\} \tag{23}$$

where P_{re} pads the image with repeated instances of the edge pixel values and A is the partial convolution matrix [27] formed by the mean filter of size $[4k+1, 4k+1]$. Given PTV_l, and using a 2D notation for pixels, we now proceed to compute the weights which are needed for image fusion. Firstly, we define

$$lmax(u, v) = \arg\max_{1\leq l\leq L} PTV_l(u, v), \tag{24}$$

$$wmax(u, v) = PTV_{lmax(u,v)}(u, v). \tag{25}$$

Secondly, we set the weights to

$$\hat{w}_l(u, v) = \begin{cases} wmax(u, v) * \kappa, & if\ l = lmax(u, v) \\ PTV_l(u, v), & otherwise \end{cases}$$

where $\kappa \in \mathbb{R}^+$ is selected to achieve a trade-off between image smoothness and sharpness [4]. By default we use $\kappa = 100$. Weights are then normalized using

$$w_l(u, v) = \frac{\hat{w}_l(u, v)}{\sum_k \hat{w}_k(u, v)}. \tag{26}$$

Finally, the initial image estimate f_0 is the weighted average of z-stack images, that is,

$$f_0 = P_{re}\sum_{l=1}^{L} w_l \circ g_l, \tag{27}$$

where the repeated boundary condition P_{re} [27] is used for padding.

The fused image f_0 combines nearly all the sharpest regions in the z-stack but it exhibits blocking artifacts caused by the fusion. When the z-stack is undersampled, the quality of f_0 drops and defocus blur remains in f_0. Overall, the fused image can be considered a good initial image estimate for PSFs estimation, since no much computation will be needed to turn it into a sharp image.

2) Sampling Distances and Depth Map Initialization: Since we do not know much about the z-stack acquisition, we initialize $\{z_l\}$ using a user provided constant ζ (e.g., an estimate of the mean $\zeta \approx \sum_{l=1}^{L} z_l/L$), that is,

$$z_l = \zeta, \quad 2 \leq l \leq L, \tag{28}$$

and the depth map using the constant ζ, i.e., $d(u, v) = \zeta$.

V. Experiments

In this section we perform three sets of experiments. We first validate the efficiency of the proposed ASGDA and compare it to the Limited memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) method [39]. Then we test our method to estimate the all-in-focus image together with both the depth map and sampling distances on synthetic data and compare it to the current state-of-the-art methods. Finally, we conduct experiments on real images. In all experiments, we set $\eta_0 = 0.1$, $tol = 1e-3$, $\lambda_f = 1e-4$ and $\kappa = 100$.

A. ASGDA Evaluation

In this section we evaluate the performance of Alg. 1 to efficiently calculate the depth map and sampling distances. The 256×256 *cameraman* is selected as the groundtruth image. To create the depth map we use a Gaussian function with $mean = [128.5, 128.5]$ and $std = 128/3$ as the normalized depth map d_n. The ground truth depth map is then set to $3 * d_n$. The sampling distances are $0, 1,$ and 2. We set $k = 4$ to define the blur size in Eq. (8), and then we generate a sequence of 3 partially in focus images to which white Gaussian noise with $std = 0.2\%$ is added, see Fig. 1(a)-(c), respectively. We use the groundtruth image as the latent image estimate, set $\lambda_d = 2e-3$ and $\zeta = 1.5$, and compare the performance of the proposed ASGDA and L-BFGS. As shown in Fig. 1, the recovered images are of similar visual quality. The one obtained by our method is, however, slightly better in PSNR terms (33.77 dB vs 33.45 dB). See the the restorations in Figs. 1(e) and 1(f).

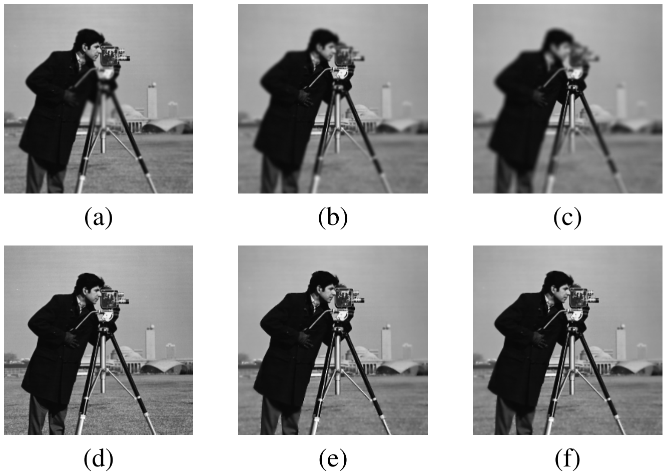


Fig. 1. ASGDA vs L-BFGS. (a)-(c) Z-stack images. (d) Groundtruth. (e) Restoration with the PSFs estimated by L-BFGS, PSNR = 33.45 dB. (f) Restoration with the PSFs estimated by ASGDA, PSNR = 33.77 dB. The restoration obtained with the PSFs estimated by ASGDA with scaling factor one is poor (PSNR = 30.80 dB) and it is not shown here.

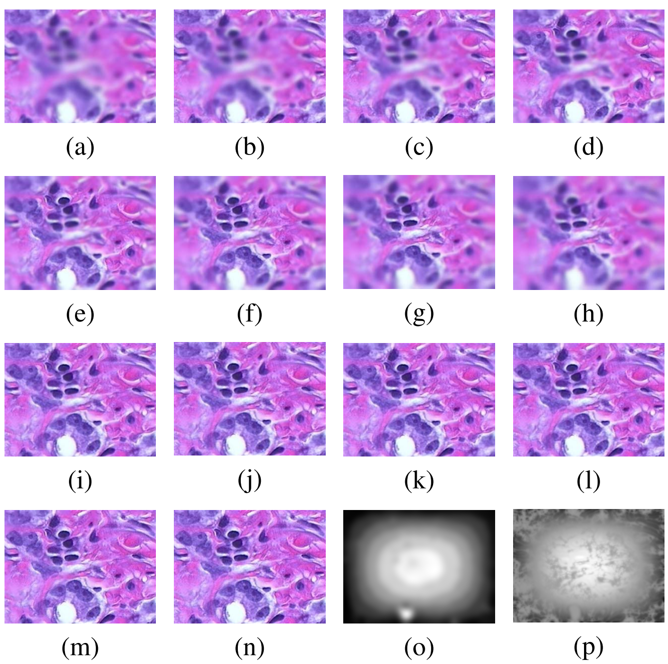


Fig. 2. Simulation results with the image formation model in Eq. (2) [12]. (a)-(h) Z-stack images synthesized using EDF [12]. (i) EDF [12] restoration, PSNR = 37.86 dB. (j) GFF [6] restoration, PSNR = 39.20 dB. (k) DSIFT [8] restoration, PSNR = 38.51 dB. (l) CNN [10] restoration, PSNR = 38.54 dB. (m) Our restoration, PSNR = 41.68 dB. (n) Groundtruth. (o) EDF depth map. (p) Our depth map. Our restoration is sharper than the others. Please zoom in for details.

Table I shows the $J_{dz}(d, \{z_l\})$ (see Eq. (15)) values obtained by both algorithms at different iterations. ASGDA drops the cost from 74.87 to 23.72 in ten iterations, whereas L-BFGS needs 20 iterations to reach a cost of 27.38. In addition, without scaling the gradient $\nabla_d J(d, \{z_l\})$ (fixing $scale = 1$ in Alg. 1), the cost decreases slowly to 48.37 after 99 iterations. This demonstrates the importance of adaptively amplifying the gradient $\nabla_d J(d, \{z_l\})$. Overall, as shown in Table I, the proposed ASGDA yields slightly more accurate sampling distances than L-BFGS and it is less time consuming.

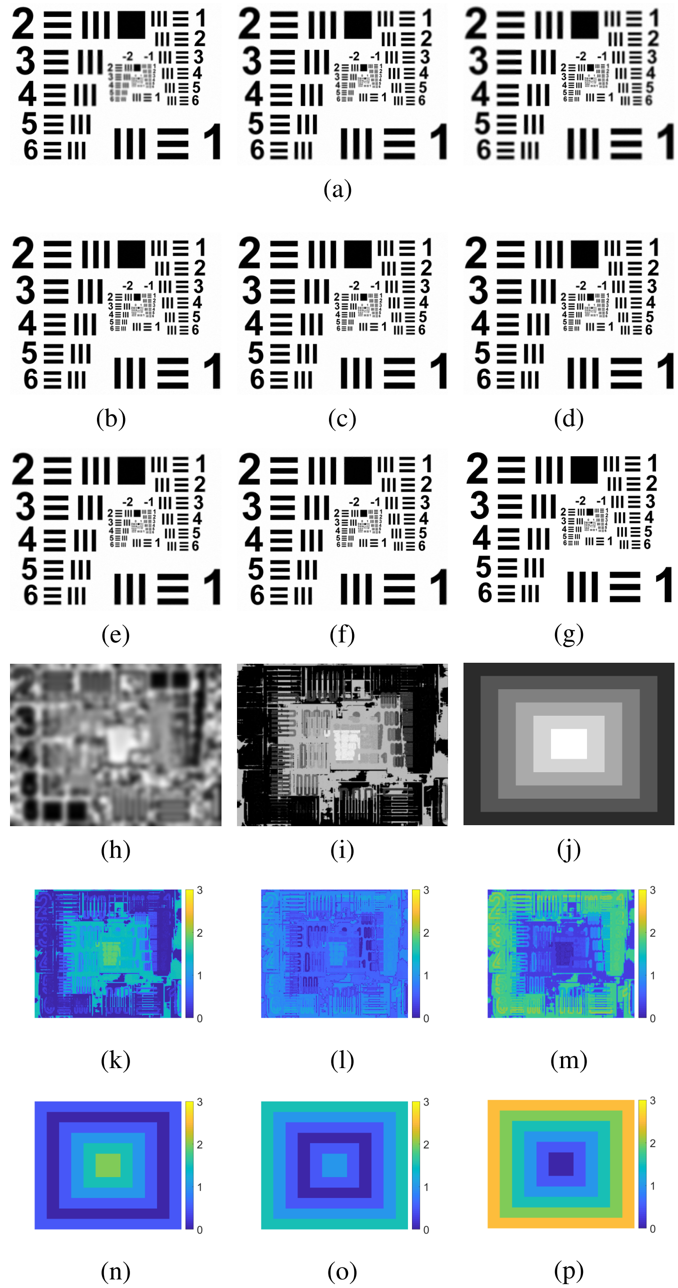


Fig. 3. Robustness against noise. (a) Z-stack with 1% Gaussian noise. (b) DSIFT restoration, PSNR = 30.64 dB. (c) CNN restoration, PSNR = 30.68 dB. (d) GFF restoration, PSNR = 30.18 dB. (e) EDF restoration, PSNR = 26.68 dB. (f) Our restoration, PSNR = 33.37 dB. (g) Groundtruth image. (h) EDF depth map. (i) Our depth map. (j) Groundtruth depth map. (k) Our heatmap $|d - z_1|$. (l) Our heatmap $|d - z_2|$. (m) Our heatmap $|d - z_3|$. (n) Groundtruth heatmap $|d^* - z_1^*|$. (o) Groundtruth heatmap $|d^* - z_2^*|$. (p) Groundtruth heatmap $|d^* - z_3^*|$. All depth maps are divided by their maximum value for visualization. Please zoom in for details.

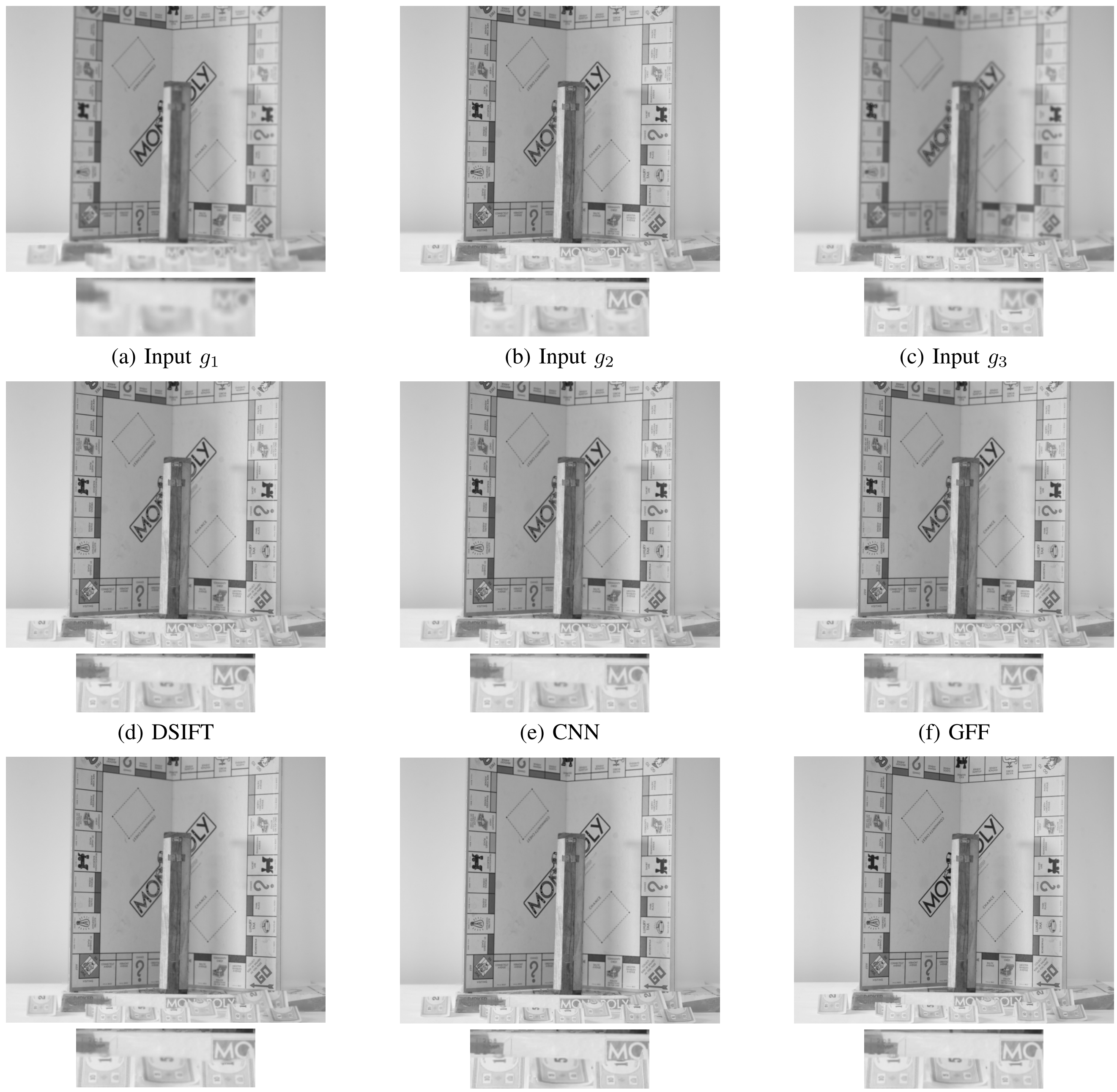
B. Synthetic Data

In this section, we compare our method, using synthetically generated z-stacks, with the deconvolution based method EDF[1] [12] and the state-of-the-art fusion based methods GFF[2] [6], DSIFT[3] [8] and CNN[3] [10]. Notice that EDF [12]

[1] Software available at http://bigwww.epfl.ch/demo/edf/

[2] MATLAB code available at http://xudongkang.weebly.com

[3] MATLAB code available at http://www.escience.cn/people/liuyu1/



(a) Input g_1 (b) Input g_2 (c) Input g_3

(d) DSIFT (e) CNN (f) GFF

(g) f_0 (h) Ours (i) Groundtruth

Fig. 4. Recovered latent sharp images from the z-stack ‘Monopoly’ (ID=16). Close-up views are also provided. PSNR values are shown in Table II (ID=16).

provides a depth map estimate, whereas the fusion methods [6], [8], [10] cannot.

1) Analyzing the Robustness of the Proposed Model: We first check the robustness of our z-stack model formulation (see Eq. (8)) when the image formation model in [12], that is, Eq. (2), is used to generate the observed images.

Since we are interested in cell pathology, the *Hestain* image is used as the latent sharp image. By setting $\eta_0 = 0.1$ and $\eta_1 = 1$ and selecting ‘dome’ as the groundtruth depth map, the EDF java software provided by Auget *et al.* [12] generates a sequence of eight partially in focus images, see Fig. 2(a)-(h). Our estimated all-in-focus image, presented Fig. 2(m), has slightly sharper textures than its competitors and leads PSNR values by $2 \sim 3$ dBs. The depth map estimated by the proposed method shows an overall dome like shape, but suffers from noticeable chaotic structures due to the ill-posed nature of flat regions. Notice that these structures in flat regions do not prevent our method from providing a good image estimate since the residuals $g_l(u,v) - [H_l f]_{(u,v)}$ in flat regions are always close to zero.

Let us now analyze the robustness of the proposed method against noise. The *USAF resolution test chart* was chosen as the latent sharp image and a pyramid of 6 levels was used as depth map, see Fig. 3(j). The pyramid depth map is in the range [0, 3] and the sampling distances are 1, 2, and 3. To each image 1% white Gaussian noise is added. Since only three of

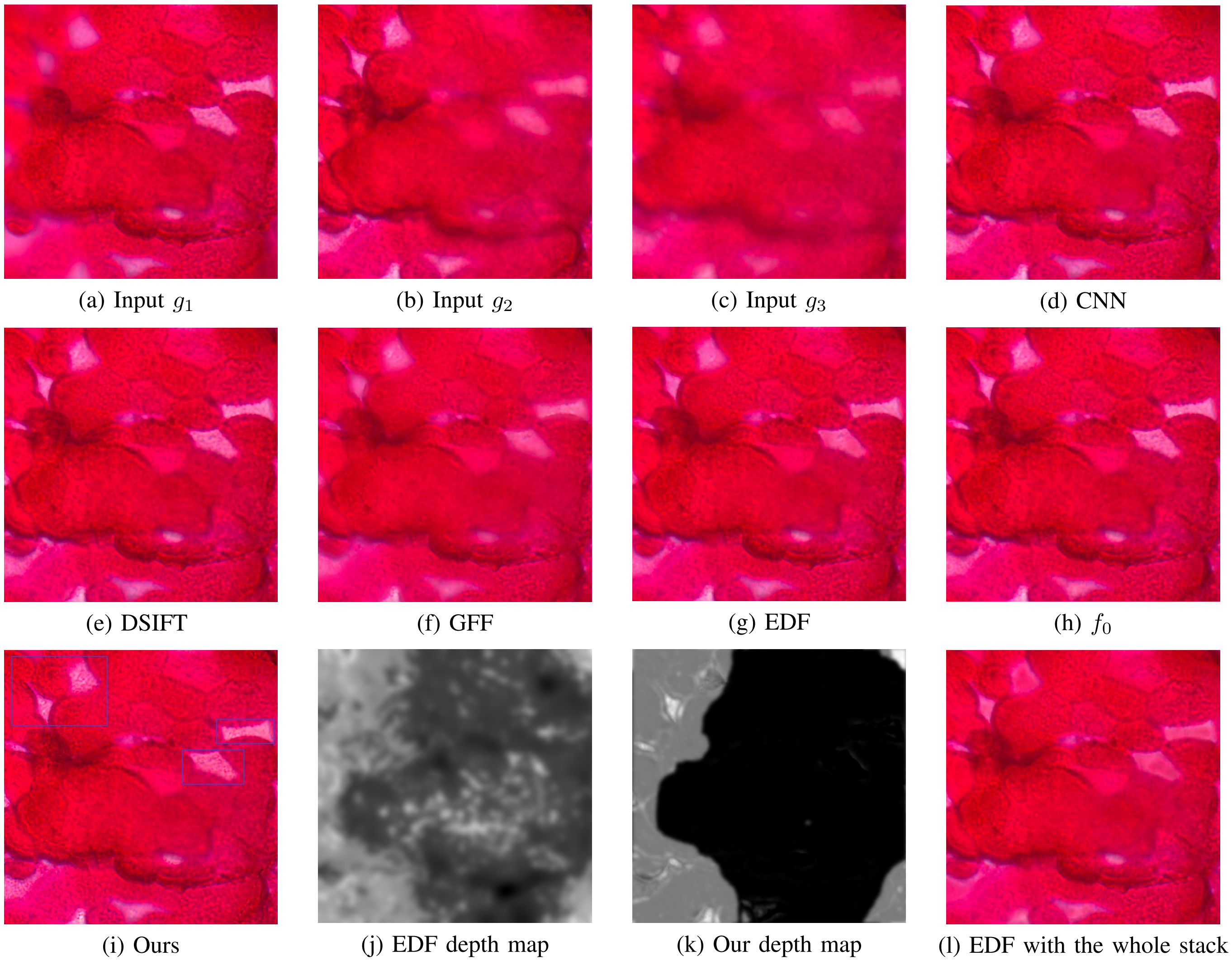


Fig. 5. All-in-focus image reconstruction from the dome z-stack provided by Auget *et al.* [12]. Our image estimate exhibits more image details and sharp edges. Please zoom in for more details.

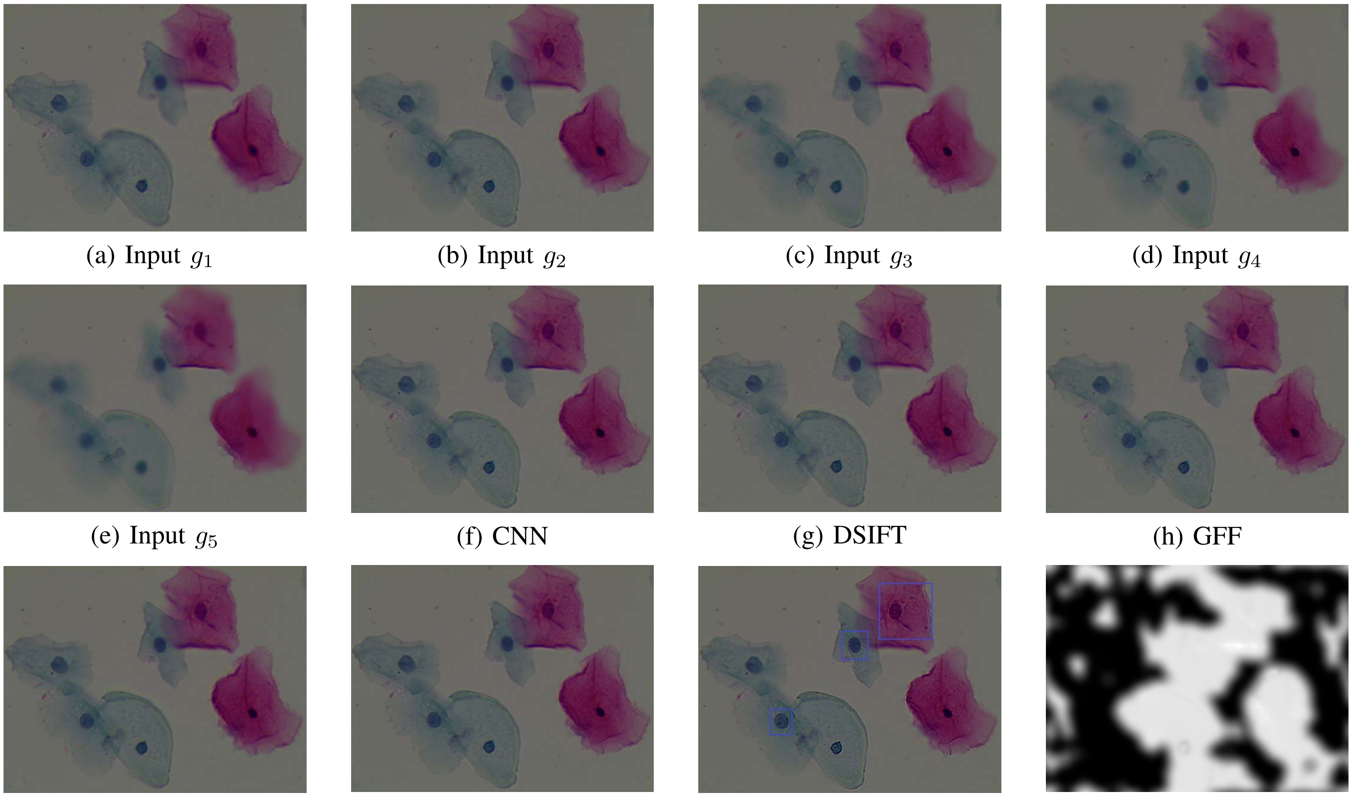
TABLE I
COMPARISON OF ASGDA AND L-BFGS

	ASGDA	ASGDA (scale=1)	L-BFGS	
Iteration	Cost	Cost	Iteration	Cost
0	74.87	74.87	0	74.87
1	62.59	56.29	1	54.36
10	23.72	51.06	5	51.22
20	20.74	50.69	10	38.07
40	16.74	50.00	20	27.38
60	14.82	49.38	30	18.84
80	13.34	48.85	40	16.33
99	12.24	48.37	49	14.20
$\{z_l\}$	[0,1.18,2.17]	[0,1.21,2.22]	$\{z_l\}$	[0,1.20,2.17]
True $\{z_l\}$	[0, 1, 2]			
Time	75.30 s	76.14s	Time	87.48 s

the six pyramid layers are sampled, more than half the area of the z-stack shown in Fig. 3(a)-(c) is out-of-focus. As we can see in the restorations in Fig. 3, the proposed method, thanks to the sparsity promoting regularization, recovers sharper image edges without amplifying the noise. The depth map estimated by EDF is quite blurry, whereas our depth map recovers parts of the pyramidal structure. Since the PSF is determined by the difference between depth map and sampling distance, we compare the PSFs by visualizing the heatmap $|d - z|$ and plot in Fig. 3(k)-(p) our estimated $|d - z_l|$ and groundtruth $|d^* - z_l^*|$.

2) Comparison With the State-of-the-Art Fusion Methods: The *scenes2006* half-size dataset[4] [38], which contains 21 color image and depth map pairs, is downloaded. All color images are converted to gray ones. The depth maps are linearly re-scaled to the interval $[0, 5]$ and the sampling distances are set to $5/3$, $10/3$, and 5. In summary, we obtain 21 z-stacks. A 0.2% white Gaussian noise is added to each individual image. The PSNR values obtained by the 5 methods are presented in Table II. As we can see, the 4 image fusion methods (f_0 stands for the result of the PTV based fusion method) perform similarly in PSNR terms, whereas the proposed method achieves much higher average PSNR than its competitors. It outperforms the best fusion method (DSIFT) by more than 1 dB.

[4]Available at http://vision.middlebury.edu/stereo/data/scenes2006/



(a) Input g_1 (b) Input g_2 (c) Input g_3 (d) Input g_4

(e) Input g_5 (f) CNN (g) DSIFT (h) GFF

(i) EDF (j) f_0 (k) Ours (l) Our depth map

Fig. 6. Comparison of image estimation from a z-stack of cervical cells. The proposed method recovers sharper image edges and more tiny details.

TABLE II
PSNR COMPARISON OF FUSION METHODS ON DATASET [38]

ID	DSIFT	CNN	GFF	f_0	Ours
1	36.22	36.27	36.27	36.17	36.75
2	41.69	41.64	41.89	41.33	41.75
3	41.68	41.36	41.53	40.83	43.31
4	37.62	37.25	37.65	37.58	38.72
5	39.45	39.54	39.47	39.48	39.92
6	39.37	37.87	38.66	38.94	39.86
7	37.77	37.72	37.74	37.67	40.01
8	38.61	38.51	38.53	38.47	40.18
9	37.32	36.83	36.83	37.26	38.68
10	36.18	35.92	36.03	35.99	37.16
11	36.10	35.32	35.65	35.82	36.47
12	41.15	40.49	40.39	40.88	41.97
13	43.50	43.58	43.51	43.30	44.60
14	38.97	38.83	39.16	38.66	41.92
15	38.04	37.91	37.84	37.48	39.89
16	35.95	35.65	35.71	35.71	37.78
17	47.30	47.65	47.45	46.76	48.17
18	37.24	36.75	36.76	36.97	38.74
19	36.63	36.36	36.65	36.54	38.24
20	43.15	42.85	42.77	43.00	44.23
21	41.47	42.09	42.22	41.36	41.67
Avg	39.30	39.07	39.18	39.06	40.48

Figure 4 shows an example for visual comparison. As shown in the close-up views in Fig. 4(a)-(c), the digit *5* at the bottom is blurred in all the acquired images because the in focus position is between two optical sections and the fusion method cannot remove such blur. The digit *5* recovered by our method, which is shown in Fig. 4(h), is as sharp as the one in the groundtruth image, see Fig. 4(i).

TABLE III
RELATIVE ℓ_2-NORM ERRORS OF THE ESTIMATED SAMPLING DISTANCES

ID	Relative error	ID	Relative error
1	0.0194	12	0.0650
2	0.7172	13	0.1497
3	0.0491	14	0.0478
4	0.0969	15	0.0688
5	0.1295	16	0.0749
6	0.0952	17	0.7000
7	0.1476	18	0.0847
8	0.0544	19	0.1325
9	0.0840	20	0.1261
10	0.1220	21	0.7665
11	0.4476		

To evaluate the estimated sampling distances we utilize the relative ℓ_2-norm error, defined as $\sqrt{\sum_{i=1}^{3}(z_i - z_i^*)^2 / \sum_{i=1}^{3} z_i^{*2}}$. For calculating the relative ℓ_2-norm error, $z_1^* = 0, z_2^* = 5/3$, and $z_3^* = 10/3$ are the right groundtruth sampling distances. Table III shows them. The four largest relative $\ell_2 - norm$ errors: 0.72, 0.45, 0.7 and 0.76 correspond to images 2, 11, 17, and 21, respectively. Notice that these errors are far away from the rest. Interestingly, although the estimated sampling distances are very inaccurate, e.g., images 11 and 17 in Table III, the PSNRs of the recovered images are better than those of the state-of-the-art fusion methods.

C. Real Data

We first evaluate the performance of the proposed method on the 'dome' z-stack provided by Auget *et al.* [12].

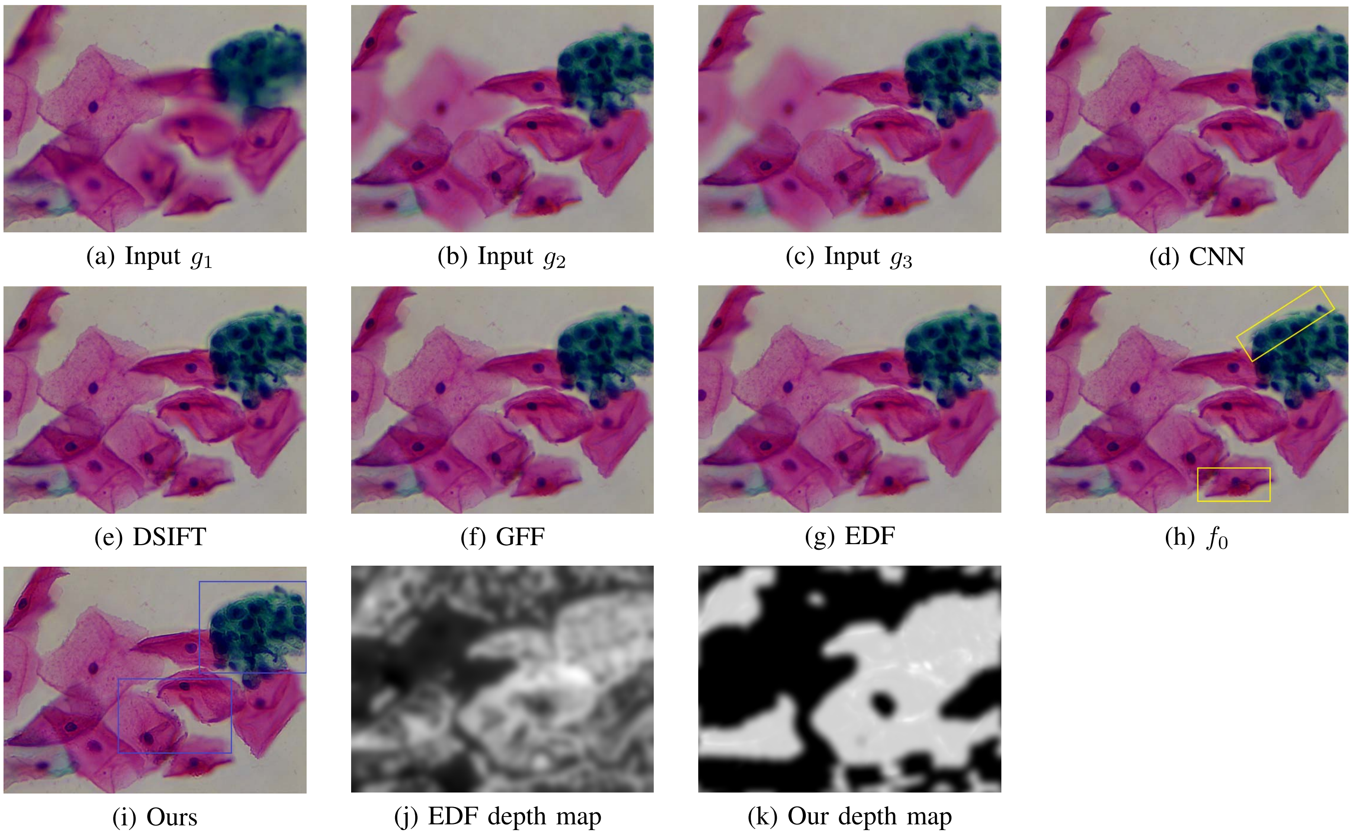


Fig. 7. Image reconstruction from a undersampled z-stack of overlapping cervical cells. Our image estimate exhibits much less fusion artifacts and more sharp nucleus and image details. Please zoom in for more details.

The 'dome' z-stack consists of 20 1450 $\times$ 1996 color images. Estimating the all-in-focus image from the whole sequence requires more than one TB of memory to store the PSFs only. Instead we only use the first three images and crop out a 600 $\times$ 600 square region, where the first three images are partially in focus but the remaining 17 images are entirely out-of-focus and hence useless for image restoration.

The cropped 600 $\times$ 600 subimages are presented in Fig. 5(a)-(c), they exhibit blur where the in focus position falls between two optical sections. From the image estimates shown in Fig. 5(d)-(i), it is clear that the proposed method significantly removes the blur and recovers more tiny textures and sharper image edges, see the blue boxes marked in Fig. 5(i). The recovered depth map in Fig. 5(k) shows that the dark and gray regions match the in-focus regions of g_1 and g_2. Furthermore, the very bright top-right corner matches the in-focus region of input g_3. We also present in Fig. 5(l) the image estimated by EDF using the whole z-stack, which is not better than our recovered image.

We now show the performance of the proposed method on two additional real z-stacks of cervical cell images of dimension 384 $\times$ 512. The first z-stack, presented in Fig. 6(a)-(e), is sampled nearly evenly along the optical axis and consists of 5 cervical cell images with some overlapping. The image estimates shown in Fig. 6(f)-(k) demonstrate that the proposed method recovers more details such as tiny nucleolus and fine cell structures, see the blue boxes marked in Fig. 6(k).

The second z-stack of cervical cell images, which is sampled unevenly along the optical axis, is shown in Fig. 7(a)-(c). This is a very challenging image set since: a) the sequence is undersampled and b) a number of cells overlap. As we can see in Fig. 7, the nucleus in the blue boxes in Fig. 7(i) are sharp while, at the same time, they are blurry in all individual acquisitions in Fig. 7(a)-(c). The other methods are not capable of removing the blur in Fig. 7(a)-(c), whereas the proposed method makes all the nucleus sharper beyond the input z-stack and recovers more fine cell structures. Moreover, the fusion artifacts (e.g., around the nucleus in the yellow boxes of Fig. 7(h)) caused by PTV are reduced by our method, as shown in Fig. 7(i). In addition, as we can see in Fig. 7(j) and Fig. 7(k), the depth map estimated by our proposed method better matches the input z-stack and exhibits much less noise than the depth map estimated by EDF. This experiment shows that the proposed method not only reduces fusion artifacts, but also recovers sharper nucleus and image edges.

VI. Conclusion

In this work, we have proposed a parameter free Gaussian PSF model for extended depth of field in brightfield microscope. This recovery problem, which is extremely ill-posed when the z-stack is undersampled, has been tackled here using a maximum a posteriori formulation which assumes a Gaussian prior for the depth map and a sparse prior for high-pass filtered versions of the latent sharp image. For depth and sampling distance estimation, we propose an adaptively scaled gradient decent algorithm. Numerical results demonstrate its superior performance against L-BFGS. Results on synthetic and real data show that, in most of cases, the proposed method estimates the sampling distances with low relative error and recovers a smooth depth map and a latent all-in-focus image with sharper edges and less fusion artifacts. A MATLAB implementation of this work and the data are available at http://site.google.com/view/cfedf.

References

[1] A. P. Pentland, "A new sense for depth of field," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-9, no. 4, pp. 523–531, Jul. 1987.

[2] S. K. Nayar and Y. Nakagawa, "Shape from focus," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 8, pp. 824–831, Aug. 1994.

[3] A. N. Rajagopalan and S. Chaudhuri, "An MRF model-based approach to simultaneous recovery of depth and restoration from defocused images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 7, pp. 577–589, Jul. 1999.

[4] S. Pertuz, D. Puig, M. A. Garcia, and A. Fusiello, "Generation of all-in-focus images by noise-robust selective fusion of limited depth-of-field images," *IEEE Trans. Image Process.*, vol. 22, no. 3, pp. 1242–1251, Mar. 2013.

[5] S. Pertuz, D. Puig, and M. A. Garcia, "Analysis of focus measure operators for shape-from-focus," *Pattern Recognit.*, vol. 46, no. 5, pp. 1415–1432, 2013.

[6] S. Li, X. Kang, and J. Hu, "Image fusion with guided filtering," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2864–2875, Jul. 2013.

[7] M. Nejati, S. Samavi, and S. Shirani, "Multi-focus image fusion using dictionary-based sparse representation," *Inf. Fusion*, vol. 25, pp. 72–84, Sep. 2015.

[8] Y. Liu, S. Liu, and Z. Wang, "Multi-focus image fusion with dense SIFT," *Inf. Fusion*, vol. 23, pp. 139–155, May 2015.

[9] G. Wang, W. Li, X. Yin, and H. Yang, "All-in-focus with directional-max-gradient flow and labeled iterative depth propagation," *Pattern Recognit.*, vol. 77, pp. 173–187, May 2018.

[10] Y. Liu, X. Chen, H. Peng, and Z. F. Wang, "Multi-focus image fusion with a deep convolutional neural network," *Inf. Fusion*, vol. 36, pp. 191–207, Jul. 2017.

[11] H. Tang, B. Xiao, W. Li, and G. Wang, "Pixel convolutional neural network for multi-focus image fusion," *Inf. Sci.*, vols. 433–434, pp. 125–141, Apr. 2017.

[12] F. Aguet, D. Van De Ville, and M. Unser, "Model-based 2.5-D deconvolution for extended depth of field in brightfield microscopy," *IEEE Trans. Image Process.*, vol. 17, no. 7, pp. 1144–1153, Jul. 2008.

[13] X. Lin, J. Suo, X. Cao, and Q. Dai, "Iterative feedback estimation of depth and radiance from defocused images," in *Proc. ACCV*, 2013, pp. 95–109.

[14] B. Meher, S. Agrawal, R. Panda, and A. Abraham, "A survey on region based image fusion methods," *Inf. Fusion*, vol. 48, pp. 119–132, Aug. 2019.

[15] M. S. Muhammad and T.-S. Choi, "Sampling for shape from focus in optical microscopy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 3, pp. 564–573, Mar. 2012.

[16] P. Favaro, S. Soatto, M. Burger, and S. J. Osher, "Shape from defocus via diffusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 3, pp. 518–531, Mar. 2008.

[17] A. Levin, R. Fergus, F. Durand, and W. T. Freeman, "Image and depth from a conventional camera with a coded aperture," *ACM Trans. Graph.*, vol. 26, no. 3, p. 70, 2007.

[18] J. Shi, L. Xu, and J. Jia, "Just noticeable defocus blur detection and estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 657–665.

[19] L. D'Andrès, J. Salvador, A. Kochale, and S. Süsstrunk, "Non-parametric blur map regression for depth of field extension," *IEEE Trans. Image Process.*, vol. 25, no. 4, pp. 1660–1673, Apr. 2016.

[20] A. Shajkofci and M. Liebling, "Semi-blind spatially-variant deconvolution in optical microscopy with local point spread function estimation by use of convolutional neural networks," in *Proc. ICIP*, Oct. 2018, pp. 3818–3822.

[21] M. Lee and Y. W. Tai, "Robust all-in-focus super-resolution for focal stack photography," *IEEE Trans. Image Process.*, vol. 25, no. 4, pp. 1887–1897, Apr. 2016.

[22] E. R. Dowski and W. T. Cathey, "Extended depth of field through wave-front coding," *Appl. Opt.*, vol. 34, no. 11, pp. 1859–1866, 1995.

[23] C. Pan, J. Chen, R. Zhang, and S. Zhuang, "Extension ratio of depth of field by wavefront coding method," *Opt. Express*, vol. 16, no. 17, pp. 13364–13371, Aug. 2008.

[24] R. N. Zahreddine and C. J. Cogswell, "Total variation regularized deconvolution for extended depth of field microscopy," *Appl. Opt.*, vol. 54, no. 9, pp. 2244–2254, Mar. 2015.

[25] L. Sheng and H. Hong, "Extended depth-of-field microscopic imaging with a variable focus microscope objective," *Opt. Express*, vol. 19, no. 1, p. 353, 2011.

[26] P. Favaro, "Recovering thin structures via nonlocal-means regularization with application to depth from defocus," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 1133–1140.

[27] X. Zhou, F. Zhou, X. Bai, and B. Xue, "A boundary condition based deconvolution framework for image deblurring," *J. Comput. Appl. Math.*, vol. 261, pp. 14–29, May 2014.

[28] S. Pertuz, M. A. Garcia, and D. Puig, "Efficient focus sampling through depth-of-field calibration," *Int. J. Comput. Vis.*, vol. 112, no. 3, pp. 342–353, 2015.

[29] H. H. Hopkins, "The frequency response of a defocused optical system," *Proc. Roy. Soc. Math. Phys. Eng. Sci.*, vol. 231, no. 1184, pp. 91–103, 1955.

[30] P. A. Stokseth, "Properties of a defocused optical system," *J. Opt. Soc. Amer.*, vol. 59, no. 10, pp. 1314–1321, 1969.

[31] S. Pertuz, M. A. Garcia, D. Puig, and H. Arguello, "A closed-form focus profile model for conventional digital cameras," *Int. J. Comput. Vis.*, vol. 124, no. 3, pp. 273–286, Sep. 2017.

[32] X. Zhou, J. Mateos, F. Zhou, R. Molina, and A. K. Katsaggelos, "Variational Dirichlet blur Kernel estimation," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5127–5139, Dec. 2015.

[33] X. Zhou, M. Vega, F. Zhou, R. Molina, and A. K. Katsaggelos, "Fast Bayesian blind deconvolution with huber super Gaussian priors," *Digit. Signal Process.*, vol. 60, pp. 122–133, Jan. 2017.

[34] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D, Nonlinear Phenomena*, vol. 60, nos. 1–4, pp. 259–268, 1992.

[35] J. Nocedal and S. J. Wright, *Numerical Optimization*. New York, NY, USA: Springer, 1999.

[36] X. Zhou, R. Molina, F. Zhou, and A. K. Katsaggelos, "Fast iteratively reweighted least squares for L_p regularized image deconvolution and reconstruction," in *Proc. ICIP*, Oct. 2014, pp. 1783–1787.

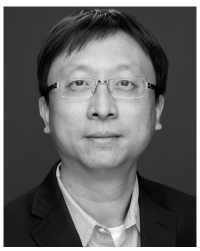
[37] D. Krishnan, T. Tay, and R. Fergus, "Blind deconvolution using a normalized sparsity measure," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2011, pp. 233–240.

[38] D. Scharstein and C. Pal, "Learning conditional random fields for stereo," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.

[39] D. C. Liu and J. Nocedal, "On the limited memory BFGS method for large scale optimization," *Math. Program.*, vol. 45, no. 1, pp. 503–528, 1989.

Xu Zhou received the B.S. degree in applied mathematics from Central South University in 2009 and the Ph.D. degree in pattern recognition and intelligent system from Beihang University in 2016. From 2013 to 2014, he was a Visiting Student with Visual Information Processing Group, Department of Computer Science and Artificial Intelligence, University of Granada. From 2016 to 2018, he was a Postdoctoral Fellow with the School of Information Science and Technology, ShanghaiTech University, Shanghai, China. Since 2018, he has been a Postdoctoral Fellow with the School of Biomedical Engineering, Shenzhen University, Shenzhen, China. His research interests include image restoration, biomedical image registration and segmentation, machine learning, and nonconvex optimization.

Rafael Molina received the degree in mathematics (statistics) and the Ph.D. degree in optimal design in linear models from the University of Granada, Granada, Spain, in 1979 and 1983, respectively. He was the Dean of the Computer Engineering School, University of Granada, from 1992 to 2002, where he became a Professor of computer science and artificial intelligence in 2000. He was the Head of the Computer Science and Artificial Intelligence Department, University of Granada, from 2005 to 2007. He has coauthored an article that received the runner-up prize at reception for early stage researchers at the House of Commons in 2017. He has coauthored an awarded Best Student Paper at the IEEE International Conference on Image Processing in 2007, the ISPA Best Paper in 2009, and the EUSIPCO 2013 Best Student Paper. His research interest focuses mainly on using Bayesian modeling and inference in image restoration (applications to astronomy and medicine), super-resolution of images and video, blind deconvolution, computational photography, source recovery in medicine, compressive sensing, low-rank matrix decomposition, active learning, fusion, supervised learning, and crowd sourcing. He has served as an Associate Editor for *Applied Signal Processing* from 2005 to 2007 and the IEEE TRANSACTIONS ON IMAGE PROCESSING from 2010 to 2014. He has been serving as an Area Editor for *Digital Signal Processing* since 2011.



Yi Ma received the bachelor's degree in automation and applied mathematics from Tsinghua University, Beijing, China, in 1995, and the master's degree in EECS and mathematics and the Ph.D. degree in EECS from the University of California at Berkeley, Berkeley, in 2000. From 2000 to 2011, he has served on the Faculty of the ECE Department, University of Illinois at Urbana–Champaign. From 2009 to 2014, he has served as a Principal Researcher and a Research Manager for the Visual Computing Group, Microsoft Research Asia, Beijing, China. From 2014 to 2017, he was a Professor and the Executive Dean of the School of Information Science and Technology, ShanghaiTech University, China. Since January 2018, he has been with the Faculty of the EECS Department, University of California at Berkeley. He has written two textbooks *An Invitation to 3D Vision* (Springer) and *Generalized Principal Component Analysis* (Springer). He is a fellow of ACM. He has received best paper awards from ICCV, ECCV, and ACCV. He received the CAREER Award from the NSF and the YIP Award from the ONR. He has served as an Associate Editor for IJCV, SIIMS, SIMODS, the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE (PAMI), and the IEEE TRANSACTIONS ON INFORMATION THEORY.

Tianfu Wang received the Ph.D. degree in biomedical engineering from Sichuan University in 1997. He is currently a Professor with the Department of Biomedical Engineering, School of Medicine, Shenzhen University, where he is also the Associate Chair of the School of Medicine. His research interests include ultrasound image analysis, medical image processing, pattern recognition, and medical imaging.

Dong Ni received the bachelor's and master's degrees in biomedical engineering from Southeast University, Nanjing, China, in 2000 and 2003, respectively, and the Ph.D. degree in computer science and engineering from The Chinese University of Hong Kong, Hong Kong, in 2009. From 2009 to 2010, he was a Postdoctoral Fellow with the School of Medicine, University of North Carolina at Chapel Hill, USA. Since 2010, he has been with Shenzhen University, China, where he is currently a Professor and the Associate Dean of the School of Biomedical Engineering, Health Science Center. He founded the Medical Ultrasound Image Computing (MUSIC) Laboratory, Shenzhen University. His research interests include ultrasound image analysis, image guided surgery, and pattern recognition.