ABSTRACT

We present a Bayesian approach for 3D image reconstruction of an extended object imaged with multi-focus microscopy (MFM). MFM simultaneously captures multiple sub-images of different focal planes to provide 3D information of the sample. The naive method to reconstruct the object is to stack the sub-images along the \( z \)-axis, but the result suffers from poor resolution in the \( z \)-axis. The maximum a posteriori framework provides a way to reconstruct a 3D image according to its observation model and prior knowledge. It jointly estimates the 3D image and the model parameters. Experimental results with synthetic and real experimental data show that it enables the high-quality 3D reconstruction of an extended object from MFM.

Index Terms— Multi-focus microscopy, 3D image reconstruction, Bayesian, maximum a posteriori, total variation

1. INTRODUCTION

Multi-focus microscopy (MFM) was developed for fast acquisition of three-dimensional (3D) information [1]. In addition to the conventional microscopy setup, MFM includes a diffractive optical element (DOE), which splits the light from different focal planes into different positions on a camera (see Fig. 1 (a)). A captured MFM image includes \( K \times K \) sub-images (or tiles) of a focal stack as shown in Fig. 1 (b). By capturing a stack of \( K^2 \) focal planes with a single shot, MFM enables the capture of a dynamic scene in biological samples.

In [1], Abrahamsson et al. showed that MFM can be used for imaging movement of cells and tracking a single molecule. However, they do not propose an algorithm to reconstruct the original 3D object from an MFM measurement. Instead MFM sub-images are stacked and aligned to get the 3D volume. With this scheme, one can only obtain as many focal planes as the number of sub-images, and the \( z \) spacing of the reconstructed image is limited to the focal shift between two adjacent sub-images. Also, this approach suffers from severe out-of-focus blur, which is a common issue for wide-field microscopy [2, 3].

In order to overcome these limitations, Huang et al. [4] proposed a method to reconstruct the 3D volume from the measured MFM images. To deal with the ill-posedness of the problem, they developed

an optimization approach using sparsity (\( \ell_1 \)) and total variation (TV) regularizers. Yoo et al. [5] proposed an improved approach to reconstruct the 3D volume from a sequence of frames of the scene. Both methods achieved super-resolution in the \( z \)-axis for its 3D image reconstruction, increasing the number of the \( xy \) slices from 25 to 101.

The performance of the methods proposed in [4, 5] heavily depends on regularization parameters, which must be selected through an exhaustive search. The use of the Bayesian paradigm can help to mitigate this problem by automatically estimating all the unknowns of the problem from the data [6, 7]. Some examples of its applications in image processing are image restoration [8], blind image deconvolution [9], image and video super-resolution [10, 11] or light field acquisition [12].

In this work, we develop a Bayesian approach to the problem of 3D object reconstruction from MFM images. The MFM acquisition system is modeled by a likelihood function, which takes into account two different noise schemes: (1) the same noise for all the tiles, and (2) different noise for each tile. Then we use a prior model based on the TV norm, following [4, 5]. The reconstructed 3D object as well as the parameters of the model are estimated using maximum a posteriori (MAP) inference.

The two-dimensional (2D) TV regularizer was introduced in [13] and it has been widely used in image reconstruction to enforce piece-wise smoothness on the reconstructed images (see [12, 14]). In addition to [4,5], the 3D TV regularizer has been used in [15] in a tomography problem. However, to the best of our knowledge, a method for estimating the regularization parameter for the 3D TV function has not been reported in the literature. The Bayesian approach we develop allows extension of the methodology introduced in [14] to estimate the TV prior parameter in the 3D case.

In summary, the novelties of our work are: (1) a Bayesian modeling of the problem of 3D reconstruction from an MFM image, (2) a new noise model that allows different noise levels for each MFM sub-image, and (3) automatic estimation of all the noise variances as
The image acquisition process in MFM microscopy can be modeled as follows:
\[ \mathbf{g}_i = \mathbf{D} \mathbf{H}_i \mathbf{f} + \mathbf{e}_i, \quad i = 1, \ldots, K^2. \]  
(1)

In the \( i \)-th tile, we acquire a sub-image \( \mathbf{g}_i \in \mathbb{R}^{M \times 1} \) written as a column vector. The 3D object to be recovered, also written as a column vector \( \mathbf{f} \in \mathbb{R}^{N \times 1} \), is convolved with a 3D point spread function (PSF). In matrix notation, this 3D convolution is represented as \( \mathbf{H}_i \in \mathbb{R}^{N \times N} \). Note that for each tile, we have a different PSF. This is due to chromatic aberrations introduced by the DOE, which produce different blur for each tile in addition to the different focus (see [1,4] for details). Once the original 3D object is convolved with the PSF, we extract the central slice using the downsampling operator \( \mathbf{D} \in \mathbb{R}^{M \times N} \) to be captured on a 2D sensor. The observation includes an additive Gaussian noise \( \mathbf{e}_i \in \mathbb{R}^{M \times 1} \).

In this work, we consider two cases of noise: (1) all the tiles have the same noise variance \( \beta^{-1} \), and (2) each tile has a different noise variance \( \beta_i^{-1} \). Both cases lead to the following likelihood function
\[ p(\mathbf{g}|\mathbf{f},\beta) = \prod_{i=1}^{K^2} \frac{\beta_i^{M/2}}{(2\pi)^{M/2}} \exp \left\{ -\frac{\beta_i}{2} \| \mathbf{g}_i - \mathbf{D} \mathbf{H}_i \mathbf{f} \|_2^2 \right\}, \]  
(2)
where \( \mathbf{g} = [\mathbf{g}_1^T, \ldots, \mathbf{g}_{K^2}^T]^T \) and \( \beta = [\beta_1, \ldots, \beta_{K^2}]^T \). This observation model is related to the one introduced in [10], where the authors make use of the Bayesian framework to solve a super-resolution problem. However, there are differences between the two models. In our model, the downsampling matrix \( \mathbf{D} \) extracts a 2D slice of a 3D volume, while in [10] the downsampling matrix reduces the size of an image in 2D. Also, an additional warping matrix is considered in [10] to register the observed images.

3D object reconstruction from MFM is a highly undetermined problem because we recover many more slices than the number of tiles in the MFM images. Use of priors mitigates this problem, and we follow [4] and [5], and introduce a 3D TV prior
\[ p(\mathbf{f}|\alpha) = \frac{1}{Z_\alpha} \exp \left\{ -\alpha \text{TV}(\mathbf{f}) \right\}, \]  
(3)
where \( \text{TV}(\mathbf{f}) = \sum_{i=1}^{N} \sqrt{\Delta_1^2(f_1) + \Delta_2^2(f_2) + \Delta_3^2(f_3)} \) and \( Z_\alpha \) is the partition function depending on the parameter \( \alpha \). In order to estimate the parameters from the data, we need to know the partition function \( Z_\alpha = \int \exp \left\{ -\alpha \text{TV}(\mathbf{f}) \right\} d\mathbf{f} \), however, this integral is not tractable and \( Z_\alpha \) cannot be explicitly calculated. In [16], the authors approximate the partition function of the 2D TV prior. Applying the same procedure as in [16] for the 3D case, we calculate the integral
\[ \int_{\mathbb{R}^3} \exp \left\{ -\alpha \sqrt{u^2 + v^2 + w^2} \right\} d(u,v,w) = \frac{8\pi}{\alpha^3}, \]  
(4)
which leads to the following approximation of the prior distribution
\[ p(\mathbf{f}|\alpha) \propto \alpha^{N/2} \exp \left\{ -\alpha \text{TV}(\mathbf{f}) \right\}. \]  
(5)

Finally, all of the available information about the problem is summarized in the joint distribution
\[ p(\mathbf{g}, \mathbf{f}, \beta, \alpha) = p(\mathbf{g}|\mathbf{f},\beta)p(\mathbf{f}|\alpha)p(\beta)p(\alpha), \]  
(6)
where \( p(\beta) \) and \( p(\alpha) \) are non-informative flat priors.

### 3. Maximum A Posteriori Inference

In order to estimate all the unknowns of the model, we estimate the maximum of the posterior distribution subject to all the voxels of the restored 3D object having non-negative values, that is,
\[ \{ \hat{\mathbf{f}}, \hat{\beta}, \hat{\alpha} \} = \arg \max_{\mathbf{f},\beta,\alpha} p(\mathbf{f}, \beta, \alpha|\mathbf{g}) = \arg \min_{\beta \geq 0, \alpha \geq 0} -\log p(\mathbf{g}, \mathbf{f}, \beta, \alpha). \]  
(7)

Since we cannot minimize Eq. (7) with respect to all of the unknowns at the same time, we alternate the estimations of the 3D object, \( \mathbf{f} \), and the parameters, \( \beta \) and \( \alpha \), and iterate until the method converges.

Given the estimated parameters, \( \beta^{(n)} \) and \( \alpha^{(n)} \) for the \( n \)-th iteration, the 3D object, \( \hat{\mathbf{f}} \), is estimated by minimizing
\[ \mathbf{f}^{(n+1)} = \arg \min_{\hat{\mathbf{f}}} \sum_{i=1}^{K^2} \beta_i^{(n)} \| \mathbf{g}_i - \mathbf{D} \mathbf{H}_i \mathbf{f} \|_2^2 + \alpha^{(n)} \text{TV}(\mathbf{f}), \]  
(8)
which can be efficiently solved by using the TwIST algorithm introduced in [18], with the non-negativity constraint modification used in [4,5].

When we have the same variance of noise for all the tiles, we substitute \( \mathbf{f}^{(n+1)} \) for \( \mathbf{f} \) in Eq. (7). Taking derivatives with respect to \( \beta \) and equating to zero, we obtain the following update rule:
\[ \beta_i^{(n+1)} = \frac{MK^2}{\sum_{r=1}^{K^2} \| \mathbf{g}_i - \mathbf{D} \mathbf{H}_i \mathbf{f}^{(n+1)} \|_2^2}, \]  
(9)
Similarly, for the case of different noise variance for each tile, we obtain the following update rule for \( \beta \):
\[ \beta_i^{(n+1)} = \frac{M}{\| \mathbf{g}_i - \mathbf{D} \mathbf{H}_i \mathbf{f}^{(n+1)} \|_2^2}, \quad i = 1, \ldots, K^2. \]  
(10)

To update \( \alpha \), we substitute \( \mathbf{f}^{(n+1)} \) for \( \mathbf{f} \) in Eq. (7), and taking derivatives with respect to \( \alpha \) and equating to zero, we obtain the following update rule:
\[ \alpha_i^{(n+1)} = \frac{N}{2\text{TV}(\mathbf{f}^{(n+1)})}. \]  
(11)

The proposed method is summarized in Algorithm 1. Note that the objective function in Eq. (7) is always positive, and decreases in each iteration and, therefore, the convergence of the proposed algorithm is guaranteed.

**Algorithm 1 MAP-TV Reconstruction for MFM**

Require: \( n = 0 \), the observation \( \mathbf{g}, \beta^{(0)} \) and \( \alpha^{(0)} \).
1: \ repeat
2: \ Calculate \( \mathbf{f}^{(n+1)} \) using TwIST to solve problem in Eq. (8).
3: \ Calculate \( \beta^{(n+1)} \) using Eq. (9) or Eq. (10);
4: \ Calculate \( \alpha^{(n+1)} \) using Eq. (11);
5: \ \( n = n + 1; \)
6: \ until \ Convergence
4. EXPERIMENTAL RESULTS

To validate our algorithm, we performed experiments with both synthetic and real data. We compare the performance of the proposed method with the regularized least squares with TV regularizer (RLS-TV) proposed in [4]. For RLS-TV, we need to set the regularization parameter and we performed the exhaustive search for the parameter. Our method is tested in two schemes. In the first scheme, we assume $\beta_1 = \ldots = \beta_{t_2} = \beta$ and estimate $\beta$ as in Eq. (9) (MAP-TV1). MAP-TV1 assumes the same level of noise for all the tiles as RLS-TV. In the second scheme, we estimate each $\beta_i$ separately as in Eq. (10) (MAP-TV2).

4.1. Simulation with synthetic data

Our synthetic data is based on experimental data from a 3D image stack of a bacterium obtained by confocal microscopy. It is scaled, cropped, and located to the center of volume composed of $50 \times 50 \times 51$ voxels. Fig. 2 (a) shows the bacterium image. To visualize 3D image effectively, we cropped the region around the bacterium to be the size of $35 \times 35 \times 21$ and the center slices of the volume in $xy$, $yz$, and $zx$ planes are shown in top-left, top-right, and bottom-left of the figure, respectively.

To simulate the MFM experiment, we used the forward model defined in Eq. (1) with a measured PSF. We simulated two different levels of Gaussian noise, setting the standard deviation of the noise to $\sigma \in \{0.02, 0.05\}$. At the same time, we also simulated two types of noise for MFM: (1) Gaussian noise whose level is based on the maximum intensity over the whole MFM image, producing the same level of noise for all the sub-images (type 1 noise), and (2) Gaussian noise whose level is proportional to the maximum intensity of each sub-image, producing different level of noise for each sub-image (type 2 noise). By combination of two levels and two types, four different noise cases are simulated. The two cases are shown in Fig. 3. From the figures, we can see that the noise level is the same for the whole MFM image in Fig. 3 (a), while each tile has a different amount of noise in Fig. 3 (b).

The performance of reconstruction in PSNR and SSIM is summarized in Table 1. We performed the same simulations 10 times with different random number generator seeds for noise, and calculated the average of the PSNRs and SSIMs of the reconstruction. As described in the table, RLS-TV results in higher PSNRs and SSIMs in most cases. It is reasonable that RLS-TV produces better solutions than MAP-TV1 because we perform an exhaustive search for RLS-TV. The optimal value of the regularization parameter for RLS-TV is chosen based on PSNR/SSIM. MAP-TV1 jointly estimates the parameters corresponding the regularization parameters ($\alpha, \beta$) and the 3D image, based on posterior probability but it does not necessarily match the best PSNR parameter. The solution is biased toward the prior that prefers a piecewise smooth image. This argument also applies to MAP-TV2. In fact, for type 2 noise, the MAP-TV2 even produces comparable or better PSNR and SSIM. This is because RLS-TV is based on the assumption that the noise is the same for all tiles, while MAP-TV2 considers different noise levels for each tile.

The reconstructed images are shown in Fig. 2. Visual quality of the reconstructed images from MAP-TV1/2 are clearly better than those from RLS-TV. The reconstructed images from RLS-TV are distorted in $xy$ and elongated in the $z$-axis. In contrast, MAP-TV1/2 produce more smooth and natural images. The overall intensity level becomes a bit lower than the ground truth but this can be easily compensated. We also notice the different performance between MAP-TV1 and MAP-TV2. The results are similar but MAP-TV2 gives sharper images and less elongation in the $z$-axis. We observe that the difference between them becomes more obvious with type 2 noise. Because MAP-TV2 is able to estimate different noise for each tile, it also reconstructs higher-quality images.

Fig. 4 shows the performance of parameter estimation for noise. For noise type 1, both MAP-TV1 and MAP-TV2 estimate $\beta$ close to the ground truth. The estimated $\beta_i$ from MAP-TV2 is varying for each tile even though it has the fixed noise level. This is because the

Fig. 2. Image of a fluorescnetly labelled bacterium and reconstructed images from MFM measurement with noise type 2, $\sigma = 0.02$. (a) Ground truth from a confocal microscopy image stack, (b) RLS-TV, (c) MAP-TV1, and (d) MAP-TV2

Fig. 3. Simulated MFM data based on experimental data of a fluorescnetly labelled bacterium with two different types of noise. Each of the 25 tiles is at a different focal plane of the same object. (a) Type 1 noise, $\sigma_n = 0.05$, and (b) type 2 noise, $\sigma_n = 0.05$. 

Fig. 4. Shows the performance of parameter estimation for noise.
Table 1. Performance of MFM image reconstruction algorithms. Average PSNR and SSIM are calculated over 10 realizations of noise.

<table>
<thead>
<tr>
<th>Noise Type</th>
<th>PSNR</th>
<th>SSIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type 1, (\sigma = 0.02)</td>
<td>36.67</td>
<td>0.9826</td>
</tr>
<tr>
<td>Type 1, (\sigma = 0.05)</td>
<td>35.01</td>
<td>0.9647</td>
</tr>
<tr>
<td>Type 2, (\sigma = 0.02)</td>
<td>37.40</td>
<td>0.9891</td>
</tr>
<tr>
<td>Type 2, (\sigma = 0.05)</td>
<td>35.81</td>
<td>0.9811</td>
</tr>
</tbody>
</table>

Fig. 4. Estimation of \(\beta\) with different types and levels of noise. (a) Type 1 noise, \(\sigma = 0.02\), (b) type 1 noise, \(\sigma = 0.05\), (c) type 2 noise, \(\sigma = 0.02\), and (d) type 2 noise, \(\sigma = 0.05\).

captured object in the tiles affects the variance of the signal. As can be seen in Fig. 4 (c) and (d), MAP-TV2 estimates the different noise level precisely, while MAP-TV1 estimates the average noise level.

4.2. Real experiments

We captured experimental MFM images of a tumbling bacterium with our MFM system [4]. The effective pixel size is \(98\, \text{nm} \times 98\, \text{nm}\) and the pixel resolution of the camera is \(1024 \times 1024\). The image from the MFM consists of \(5 \times 5\) tiles, and 150 \times 150 pixels in each tile, which is equivalent to \(14.7\, \mu\text{m} \times 14.7\, \mu\text{m}\) in sample space. Fig. 5 shows a single MFM image from a 200 frame video.

The 3D image reconstructions from the experimental MFM data are shown in Fig. 6. We cropped the volume as described in Section 4.1 for the simulated data. Fig. 6 (a) shows the 3D volume obtained by stacking the tiles with cubic interpolation along the z-axis. It is very noisy and also blurred. Severe out-of-focus blur causes poor resolution in the z-axis. RLS-TV and MAP-TV1/2 produce high-quality reconstructions, achieving axial super-resolution, as shown in Fig. 6 (b-d). To compare these methods, MAP-TV is superior to RLS-TV in two aspects. First, it achieves better axial super-resolution. Looking at the \(xz\) and \(yz\) cuts of the 3D volume in Fig. 6, MAP-TV1/2 produces a sharper reconstruction with less elongation in the z-axis. Second, the reconstructions from MAP-TV are automatically estimated with other model parameters, while the reconstruction by RLS-TV is obtained by visual inspection after an exhaustive search.

5. CONCLUSION

We have introduced a Bayesian modeling for the problem of 3D image reconstruction from MFM images. The model we developed considers two scenarios: (1) the same noise variance for all of the tiles, and (2) different noise variance for each tile. In order to estimate the parameter of the prior distribution, we have approximated the partition function of the 3D TV prior. MAP inference has been performed to derive an iterative algorithm that automatically estimates all of the unknowns of the problem from the data by alternating between the 3D object estimation and model parameter estimation. The Bayesian modeling and MAP-TV have been evaluated using both synthetic and experimental MFM data. MAP-TV obtained high-quality reconstructions of the 3D object in both cases, including accurate estimations of the noise variances in the synthetic experiment. It has been demonstrated to be competitive comparing with RLS-TV with an exhaustive search of the model parameters.
6. REFERENCES


