A Theory of Defocus via Fourier Analysis

Paolo Favaro
Heriot-Watt University
Edinburgh, UK
p.favaro@hw.ac.uk
http://www.eps.hw.ac.uk/ pf21

Alessandro Duci
Arivis Multiple Image Tools, GmbH
Rostock, Germany
alessandro.duci@arivis.com

Abstract

In this paper we present a novel theory to analyze defocused images of a volume density by exploiting well-known results in Fourier analysis and the singular value decomposition. This analysis is fundamental in two respects: First, it gives a deep insight into the basic mechanisms of image formation of defocused images, and second, it shows how to incorporate additional a-priori knowledge about the geometry and photometry of the scene in restoration algorithms. For instance, we show that the case of a scene made of a single surface results in a simple constraint in the Fourier domain. We derive two basic types of algorithms for volumetric reconstruction: One based on a dense set of defocused images, and one based on a sparse set of defocused images. While the first one excels in simplicity, the second one is of more practical use. Both algorithms are tested on real and synthetic data.

1. Introduction

The problem of estimating the volume of a scene from its defocused images belongs to the class of inverse problems [2], and as such it can be found in different forms in several fields, including microscopy [8, 4, 1], tomography [7], and, more recently, in the field of Computer Vision [9, 3, 5]. Typically, the model of a defocused image involves a sharp image and a blur map that regulates the amount of defocus at each pixel, and that directly relates to the geometry of the scene. Because of the optical properties of lenses, the model is linear with respect to the sharp image, but nonlinear with respect to the geometry of the scene. Most methods circumvent the nonlinearity either by considering the problem locally, *i.e.*, via equifocal planes [12], so that geometry is (locally) a constant, or by designing sophisticated methods to explicitly cope with the nonlinearities [3, 5, 6].

In this paper however, we present a simple volumetric model that renders the estimation problem linear in the unknowns and, in particular, an exact convolution with a known kernel (section 3.1). The main outcome of this formulation is that one can immediately determine the maximum accuracy achievable given the camera settings, the number of pixels of each defocused image, and the number of input images. These parameters uniquely define both the Nyquist limit and a kernel that determines at what frequencies information is lost (in the idealized case of geometrical optics). The analysis of this model also shows that capturing defocused images by uniformly sampling the image plane to lens distance, might not yield the best geometry reconstructions (see [10] for analysis on the case of two images). Similarly, we will see that the accuracy of the reconstructed geometry is not uniform along the optical axis. Finally, although sensitive to noise, 3-D deconvolution can be performed using images from ordinary microscopes, making it an inexpensive alternative to laser confocal microscopy [4].

When few defocused images are available, however, the method above is no longer suitable. In this case, one needs to implicitly interpolate the input data, and recovering the interpolating function is a non-trivial problem on its own. As an alternative, we suggest a different approach that captures more precisely the exact subspaces where the solutions live. In section 3.2 we find that given a set of N defocused images at each point in the frequency domain there is an entire subspace of at least $L - N$ solutions (possibly different), where frequencies along the optical axis are discretized with L values. Clearly, the space of solutions may become unique whenever we perform reconstruction at only N frequencies or less; however, this also means that the reconstruction along the optical axis is limited to the resolution dictated by N components. Moreover, if one is made available additional information about the scene, *e.g.*, that the volume density is concentrated on a single surface, it may be possible to incorporate such constraints in the frequency domain via simple equations of the form of eq. (3.16). Intuitively, a high localization in the spatial domain is equivalent to a poor localization in the frequency domain, and, in the limit, the localization in the frequency

978-1-4244-2243-2/08/$25.00 ©2008 IEEE

domain becomes a uniform distribution which can be tested for.

In the next section we introduce our model of a defocused images, how to rectify the model so as to simplify it (section 2.2), and how this affects defocus (section 2.3). Then, we show how to express the problem as a full 3-D deconvolution (section 3.1) and how to cope with few input images (section 3.2).

2. Finite Aperture Cameras: Modeling

In this section we introduce the image formation model of a finite aperture camera under the approximation of geometric optics (section 2.1). Then, we show how to rectify the coordinates both on the image plane and in 3-D space so that Fourier analysis can be applied (section 2.2). Finally, in section 2.3 we give more insight on how the rectification of the space changes the original image formation model.

2.1. Image Formation Model

We represent the scene as a volume density $W : \mathbb{R}^3 \mapsto [0, \infty)$ that simultaneously encodes geometry and photometry. The photometry is captured by the intensity of W while the geometry is determined by the location of the intensities of W. More precisely, given a point $\mathbf{X} \doteq [x_1\ x_2\ 1]^T x_3 \in \mathbb{R}^3$ (in projective coordinates) the volume density is defined as a function $W(x_1, x_2, x_3)$ that assigns an energy value to the point $\mathbf{X}$. Now suppose that we can change the distance v of the image plane from the lens and that there is no absorption of light in the volume density, then we can explicitly write the image formation model of a defocused image $I_0 : \Omega \subset \mathbb{R}^2 \times [0, \infty) \mapsto [0, \infty)$ as

$$I_0(\mathbf{y}, v) = \int \psi(\pi_v[\mathbf{X}], \mathbf{y}, \sigma_v(x_3)) W(\mathbf{X}) d\mathbf{X} \qquad (2.1)$$

where $\psi : \mathbb{R}^2 \times \mathbb{R}^2 \times [0, \infty) \mapsto [0, \infty)$ is the point spread function (PSF) of the optical system, $\pi_v[\mathbf{X}] \doteq v[x_1\ x_2]^T$ is the projection operator of a 3-D point in space to the image plane at distance v from the lens, and $\sigma_v(x_3)$ is the PSF blurring radius relative to the depth x_3 and the image plane to lens distance v. In this paper we approximate the PSF ψ by a Gaussian function with standard deviation σ

$$\psi(\pi_v[\mathbf{X}], \mathbf{y}, \sigma) := \frac{1}{2\pi\sigma^2} e^{-\frac{\|v\mathbf{x}-\mathbf{y}\|^2}{2\sigma^2}} \qquad (2.2)$$

where we have defined $\mathbf{x} \doteq [x_1\ x_2]^T$ for simplicity. However, our derivation can be readily applied to any other PSF model. The blurring radius (the standard deviation of the PSF) σ_v can be written as

$$\sigma_v(x_3) \doteq \gamma \frac{Dv}{2} \left| \frac{1}{F} - \frac{1}{v} - \frac{1}{x_3} \right|, \qquad (2.3)$$

where γ is a calibration parameter (the unit conversion of millimeters to pixels), D is the lens diameter (or, more precisely, the effective lens aperture), F is the focal length of the lens, and we assume that both v and x_3 are greater than F.

2.2. Image Rectification

Model (2.1) can be further simplified if we consider the following simple rectification of each defocused image

$$\begin{aligned} I(\mathbf{z}, v) &\doteq v^2 I_0(v\mathbf{z}, v) \\ &= \int v^2 \psi(\pi_v[\mathbf{X}], v\mathbf{z}, \sigma_v(x_3)) W(\mathbf{X}) d\mathbf{X} \\ &= \int \frac{1}{2\pi\bar{\sigma}_v^2(x_3)} e^{-\frac{\|\mathbf{x}-\mathbf{z}\|^2}{2\bar{\sigma}_v^2(x_3)}} W(\mathbf{X}) d\mathbf{X} \end{aligned} \qquad (2.4)$$

where

$$\bar{\sigma}_v(x_3) \doteq \frac{\sigma_v(x_3)}{v} = \gamma \frac{D}{2} \left| \frac{1}{F} - \frac{1}{v} - \frac{1}{x_3} \right|. \qquad (2.5)$$

Notice that eq. (2.4) can be rewritten as

$$\boxed{I(\mathbf{z}, v) = \int \Big(\phi(\cdot, \bar{\sigma}_v(x_3)) * W(\cdot, x_3) \Big)(\mathbf{z}) dx_3} \qquad (2.6)$$

or, equivalently,

$$\boxed{\hat{I}(\mathbf{w}, v) = \int \hat{\phi}(\mathbf{f}, \bar{\sigma}_v(x_3)) \hat{W}(\mathbf{f}, x_3) dx_3} \qquad (2.7)$$

where $\phi(\mathbf{z}, \bar{\sigma}) \doteq \frac{1}{2\pi\bar{\sigma}^2} e^{-\frac{\|\mathbf{z}\|^2}{2\bar{\sigma}^2}}$ and its Fourier transform $\hat{\phi}(\mathbf{f}, x_3) \doteq \mathcal{F}[\phi(\cdot, \bar{\sigma}(x_3))](\mathbf{f}), \mathbf{f} \in \hat{\Omega} \subset \mathbb{R}^2$; we used the more convenient notation for the volume density $W(\mathbf{x}, x_3) \doteq W(\mathbf{X})$, and, correspondingly, $\hat{W}(\mathbf{f}, x_3) \doteq \mathcal{F}[W(\cdot, x_3)](\mathbf{f})$ for its Fourier transform. Eqns. (2.6) and (2.7) are our *image formation models*.

Remark 1 *In summary, to arrive at eq.* (2.6) *we have first parameterized the volume density in projective coordinates, and then we have scaled and rectified the input images based on their image plane to lens distance. These two steps allowed us to rewrite the initial image formation model eq.* (2.1) *in a very simple form, that can be expressed as the integral of a convolution. In the next sections we will see that this form has several advantages.*

2.3. Defocus in the Rectified Model

In this section we briefly analyze the geometry of the blur cones generated by the camera finite aperture after rectification. The PSF blurring radius σ_v was defined in eq. (2.3) and identifies the following focal plane in 3-D space

$$x_{3,v} \doteq \frac{Fv}{v - F}. \qquad (2.8)$$

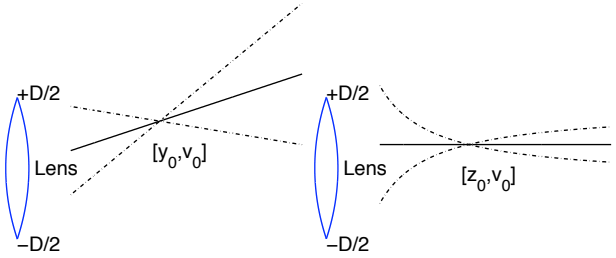

Figure 1. Left: Original blurring radii. Right: Rectified blurring radii.

The blurring radius σ_v can then be alternatively written as

$$\sigma_v(x_3) = \gamma \frac{Dv}{2} \begin{cases} \left(\frac{1}{F} - \frac{1}{v} - \frac{1}{x_3}\right) & x_3 > x_{3,v} \\ -\left(\frac{1}{F} - \frac{1}{v} - \frac{1}{x_3}\right) & F < x_3 \leq x_{3,v} \end{cases} \tag{2.9}$$

and we have explicitly separated the space into two complementary regions where the depth values x_3 have a one-to-one mapping with the blurring radii. In the case of the rectified model, the plane in focus does not change and the above equations become

$$\bar{\sigma}_v(x_3) = \gamma \frac{D}{2} \begin{cases} \left(\frac{1}{F} - \frac{1}{v} - \frac{1}{x_3}\right) & x_3 > x_{3,v} \\ -\left(\frac{1}{F} - \frac{1}{v} - \frac{1}{x_3}\right) & F < x_3 \leq x_{3,v}. \end{cases} \tag{2.10}$$

Notice that the same value of the rectified blurring radius identifies two planes in space $x_{3,\pm}(\bar{\sigma}, v)$

$$\begin{aligned} x_{3,-}(\bar{\sigma}, v) &\doteq \left(\frac{2\bar{\sigma}}{\gamma D} + \frac{1}{F} - \frac{1}{v}\right)^{-1} \\ x_{3,+}(\bar{\sigma}, v) &\doteq \left(-\frac{2\bar{\sigma}}{\gamma D} + \frac{1}{F} - \frac{1}{v}\right)^{-1}. \end{aligned} \tag{2.11}$$

Another important observation that distinguishes the original blurring radius σ_v from the rectified one, $\bar{\sigma}_v$, is that the first goes to infinity as $v \to +\infty$, while the latter converges to $\gamma \frac{D}{2} \left|\frac{1}{F} - \frac{1}{x_3}\right|$, which is a finite value. Similarly, as the 3-D point being imaged goes to infinity (*i.e.*, $x_3 \to \infty$) then $\sigma_v \to \gamma \frac{Dv}{2} \left|\frac{1}{F} - \frac{1}{v}\right|$, while $\bar{\sigma}_v \to \gamma \frac{D}{2} \left|\frac{1}{F} - \frac{1}{v}\right|$. These facts are illustrated in Figure 1.

3. Finite Aperture Cameras: Analysis

In this section we present two approaches to solve the problem of reconstructing the volume density W given several defocused images. The first approach introduces an additional change of coordinates that renders the image formation model (2.6) an exact convolution (section 3.1). This enables the use of deconvolution methods to retrieve the volume density, provided that enough images are given as input [11]. The second approach exploits the linearity of the model to obtain an efficient least-squares solution (section 3.2).

3.1. The Space of all Defocused Images and Volumetric Deconvolution

Recall the image formation model (2.6)

$$I(\mathbf{z}, v) = \int \Big(\phi(\cdot, \bar{\sigma}_v(x_3)) * W(\cdot, x_3)\Big)(\mathbf{z}) dx_3 \tag{3.1}$$

and notice that if we let $\mu \doteq \frac{1}{F} - \frac{1}{v}$, $\nu \doteq \frac{1}{x_3}$, $\tilde{W}(\mathbf{x}, \nu) \doteq \frac{1}{\nu^2} W(\mathbf{x}, \frac{1}{\nu})$, and $\tilde{\sigma}(\mu - \nu) \doteq \bar{\sigma}_{\frac{1}{1/F - \mu}}(\frac{1}{\nu})$ the above model can be written as

$$I(\mathbf{z}, \mu) = \int \phi(\mathbf{z} - \mathbf{x}, \tilde{\sigma}(\mu - \nu)) \tilde{W}(\mathbf{x}, \nu) d\mathbf{x} d\nu. \tag{3.2}$$

It is immediate to notice that eq. (3.2) is a convolution between a 3-D point spread function and the transformed volume density $\tilde{W}$ and, as such, it can be written as a product in Fourier domain with considerable computational advantages. The variable $I(\mathbf{z}, \mu)$, seen as a map in a 3-D domain, is the *space of all defocused images*, and eq. (3.2) shows how this space is mapped to the 3-D space outside the camera and vice versa. This model allows for standard Fourier analysis to determine the maximal resolution and accuracy of the reconstructed volume density given the input images. Indeed, notice that if the 3-D point spread function is not a Dirac delta, some of the content of the volume density will be irremediably lost no matter how many defocused images are used. Based on this model it is immediate to conclude that one way to obtain an invertible 3-D PSF is to have a very wide lens or very small pixels, so that the location of a point in focus can be determined very accurately (assuming a very small depth of field). Notice that the accuracy determined by this model needs to be mapped back to the un-distorted volume via $W(\mathbf{x}, x_3) = \frac{1}{x_3^2} \tilde{W}(\frac{1}{x_3})$. Also, standard deconvolution algorithms, such as the Lucy-Richardson deconvolution algorithm, can be readily applied [2].

As the reconstruction algorithm based on the above model relies on a dense set of defocused images, if we are made available only a finite set of N images, it is necessary to provide an explicit interpolation model of the form

$$I(\mathbf{z}, \mu) \doteq \sum_{i=1}^{N} I(\mathbf{z}, \mu_i) h(\mu - \mu_i) \tag{3.3}$$

where $h : \mathbb{R} \mapsto \mathbb{R}$ is an interpolating function. Notice, however, that finding the correct interpolating function h is a non trivial task. Therefore, in the next section we look for a solution that does not rely on a dense set of defocused images.

3.2. Volumetric Least Squares

Recall the image formation model (2.7) in the frequency domain

$$\hat{I}(\mathbf{f}, v) = \int \hat{\phi}(\mathbf{f}, \bar{\sigma}_v(x_3))\hat{W}(\mathbf{f}, x_3)dx_3. \tag{3.4}$$

This model can also be written as a Hermitian inner product, *i.e.*,

$$\hat{I}(\mathbf{f}, v) = \langle \hat{\phi}(\mathbf{f}, \bar{\sigma}_v(\cdot)), \hat{W}^*(\mathbf{f}, \cdot)\rangle. \tag{3.5}$$

Now, suppose that we are given N defocused images $I(\cdot, v_1), \ldots, I(\cdot, v_N)$, obtained with image plane to lens distance $v_1, \ldots, v_N$; then, at each frequency $\mathbf{f}$ we can write the following system of linear equations

$$\begin{cases} \hat{I}(\mathbf{f}, v_1) &= \langle \hat{\phi}(\mathbf{f}, \bar{\sigma}_{v_1}(\cdot)), \hat{W}^*(\mathbf{f}, \cdot)\rangle \\ \cdots & \\ \hat{I}(\mathbf{f}, v_N) &= \langle \hat{\phi}(\mathbf{f}, \bar{\sigma}_{v_N}(\cdot)), \hat{W}^*(\mathbf{f}, \cdot)\rangle. \end{cases} \tag{3.6}$$

Thanks to Plancherel's theorem (*i.e.*, the general form of Parseval's theorem), the above system is equivalent to

$$\begin{cases} \hat{I}(\mathbf{f}, v_1) &= \langle \widehat{\phi}(\mathbf{f}, \bar{\sigma}_{v_1}(\cdot)), \widehat{W}^*(\mathbf{f}, \cdot)\rangle \\ \cdots & \\ \hat{I}(\mathbf{f}, v_N) &= \langle \widehat{\phi}(\mathbf{f}, \bar{\sigma}_{v_N}(\cdot)), \widehat{W}^*(\mathbf{f}, \cdot)\rangle \end{cases} \tag{3.7}$$

where $\widehat{(\cdot)}$ denotes the Fourier transform with respect to all arguments of the function being transformed, so that $\widehat{\phi}(\mathbf{f}, \bar{\sigma}_{v_1}(f_3)) \doteq \mathcal{F}[\hat{\phi}(\mathbf{f}, \bar{\sigma}_{v_1}(\cdot))](f_3)$ and $\widehat{W}(\mathbf{f}, f_3) \doteq \left(\mathcal{F}[\hat{W}^*(\mathbf{f}, \cdot)](f_3)\right)^* = \mathcal{F}^{-1}[\hat{W}(\mathbf{f}, \cdot)](f_3)$. For notational simplicity, let us write the PSF as the N-dimensional vector

$$\widehat{\boldsymbol{\phi}}(\mathbf{f}, f_3) \doteq \begin{bmatrix} \mathcal{F}[\hat{\phi}(\mathbf{f}, \bar{\sigma}_{v_1}(\cdot))](f_3) \\ \cdots \\ \mathcal{F}[\hat{\phi}(\mathbf{f}, \bar{\sigma}_{v_N}(\cdot))](f_3) \end{bmatrix} \tag{3.8}$$

and the N-dimensional input image vector

$$\hat{\mathbf{I}}(\mathbf{f}) \doteq \begin{bmatrix} \hat{I}(\mathbf{f}, v_1) \\ \cdots \\ \hat{I}(\mathbf{f}, v_N) \end{bmatrix} \tag{3.9}$$

so that the above system of linear equations can be written very compactly as

$$\hat{\mathbf{I}}(\mathbf{f}) = \langle \widehat{\boldsymbol{\phi}}(\mathbf{f}, \cdot), \widehat{W}(\mathbf{f}, \cdot)\rangle. \tag{3.10}$$

Then, from linear algebra we now that the solutions (in a *least-square sense*) of the above system of linear equations are given by

$$\widehat{W}(\mathbf{f}, f_3) = \widehat{\boldsymbol{\phi}}^{\dagger}(\mathbf{f}, f_3)\hat{\mathbf{I}}(\mathbf{f}) + \widehat{\boldsymbol{\phi}}^{0}(\mathbf{f}, f_3)\hat{\mathbf{a}}(\mathbf{f}) \tag{3.11}$$

where $\widehat{\boldsymbol{\phi}}^{\dagger}(\mathbf{f}, f_3) \in \mathbb{C}^N$ denotes the pseudo-inverse of $\widehat{\boldsymbol{\phi}}(\mathbf{f}, f_3)$, $\widehat{\boldsymbol{\phi}}^{0}(\mathbf{f}, f_3) \in \mathbb{C}^K$ denotes the kernel of $\widehat{\boldsymbol{\phi}}(\mathbf{f}, f_3)$, with K the dimension of the kernel, and $\hat{\mathbf{a}}(\mathbf{f}) \in \mathbb{C}^K$ a vector of coefficients at each frequency $\mathbf{f}$. The dimension K depends on the discretization of f_3. If we use L elements to represent frequencies f_3, then $K \geq L - N$ as not all images may add information at all frequencies.

Eq. (3.11) explicitly represents the subspace of all volume densities that generate the input images. If no additional information is given, one can, for example, let $\hat{\mathbf{a}}(\mathbf{f}) = 0, \forall \mathbf{f} \in \hat{\Omega}$. In section 5 we show that such choice is sufficient to recover the depth map and an image in focus of the scene. Notice that, unless additional a priori information is available, we can only reconstruct as many frequencies along f_3 as the rank of $\widehat{\boldsymbol{\phi}}(\mathbf{f}, f_3)$, *i.e.*, N frequencies or less.

3.3. Geometric Priors: Surfaces versus Volumes

In some cases we may know *a-priori* that the volume density is made of surfaces, so that it can be written as

$$W(\mathbf{x}, x_3) = r(\mathbf{x})\delta(x_3 - s(\mathbf{x})) \tag{3.12}$$

where $r, s : \Omega \mapsto [0, \infty)$ are the radiance intensity and the depth map respectively, while δ denotes the Dirac delta. This constraint can also be easily formulated in the frequency domain, where we can take advantage of the fact that a Dirac delta is transformed into a phase change with constant magnitude 1. If we take the Fourier transform of W with respect to x_3, and use the notation $\hat{W}(\mathbf{x}, f_3) \doteq \mathcal{F}[W(\mathbf{x}, \cdot)](f_3)$, we have

$$\hat{W}(\mathbf{x}, f_3) = r(\mathbf{x})e^{-jf_3 s(\mathbf{x})}. \tag{3.13}$$

Notice that the absolute value of this transform is

$$\left|\hat{W}(\mathbf{x}, f_3)\right| = r(\mathbf{x}) \tag{3.14}$$

which is constant in f_3. Then, if we take the derivative of the above equation with respect to f_3 we obtain

$$\frac{\partial}{\partial f_3}\left|\hat{W}(\mathbf{x}, f_3)\right| = 0 \qquad \forall \mathbf{x}, f_3. \tag{3.15}$$

If we compute the derivative explicitly, we have

$$\begin{aligned} \frac{\partial}{\partial f_3}\left|\hat{W}(\mathbf{x}, f_3)\right| &= \frac{\partial}{\partial f_3}\sqrt{\hat{W}(\mathbf{x}, f_3)\hat{W}^*(\mathbf{x}, f_3)} \\ &= \text{Real}\left[\frac{\partial \hat{W}(\mathbf{x}, f_3)}{\partial f_3}\frac{\hat{W}^*(\mathbf{x}, f_3)}{\left|\hat{W}(\mathbf{x}, f_3)\right|}\right] \\ &= 0 \qquad \forall \mathbf{x}, f_3. \end{aligned} \tag{3.16}$$

In our experiments however, we do not find it necessary to impose the surface constraint in eq. (3.16). The above constraint is just one example of how additional a priori knowledge can be incorporated. Similar useful constraints that can be easily formulated are the non negativity of the volume density, its sparsity along the depth coordinate and the smoothness of its surfaces.

3.4. Extracting Surface and Radiance

Once the volume density W has been reconstructed by solving (3.11), we are left with the problem of extracting the radiance r and the surface s. A simple way to do so is to exploit the structure of eq. (3.12) and compute the reconstructed radiance $\tilde{r}$ as

$$\tilde{r}(\mathbf{x}) = \int W(\mathbf{x}, x_3) dx_3 \tag{3.17}$$

and the reconstructed surface $\tilde{s}$ as

$$\tilde{s}(\mathbf{x}) = \frac{\int x_3 W(\mathbf{x}, x_3) dx_3}{\tilde{r}(\mathbf{x})}. \tag{3.18}$$

Although these are very coarse estimators (these are the zero-th and first moments of the volume density along the depth coordinate), in section 5 we will see that they are fairly robust to lack of data or noise.

4. Numerical Implementation

Although the operators $\widehat{\phi}^\dagger$ can be formally written as in eq. (3.11), the evaluation of the functional SVD is not straighforward. In our implementation we simply discretize the range values of the depth coordinates, so that $\widehat{\phi}^\dagger(\mathbf{f}, \cdot, \cdot)$ is a finite real symmetric matrix (e.g., if we consider 128 depth values, then we will have a 128×128 matrix $\widehat{\phi}^\dagger(\mathbf{f}, \cdot, \cdot)$) and the discrete SVD can be computed at each frequency $\mathbf{f}$. Notice that $\widehat{\phi}^\dagger(\mathbf{f}, \cdot, \cdot)$ depends only on the magnitude of the frequency $\mathbf{f}$ (it is rotationally symmetric before and after the Fourier transform) and therefore one could discretize the magnitude $\|\mathbf{f}\|$ and use polar coordinates to save computations. Notice that this operation is done once for all given the PSF and the camera parameters γ, F, D, $\{v_1, \ldots, v_N\}$. At run-time one can reconstruct the volume density W by using eq. (3.11) and by computing the inverse Fourier transform of $\widehat{W}$. The algorithm is summarized in Table 1.

5. Experiments

In this section we test the algorithm presented in Table 1 both on synthetic (unaffected by noise or with simulated noise) and on real (affected by sensor noise) stacks of images with focus at different planes.

5.1. Synthetic Data

In all the experiments with synthetic generated volumes, we consider a thin lens with: focal length $F = 0.012m$, lens aperture $D = 0.02m$, and meter-to-pixel conversion factor γ ranging between $100pixel/m$ and $300pixel/m$. We assume a volume density W contained between $.5m$ and $.9m$ from the lens plane(coordinate x_3) and test the reconstruction performances in two cases:

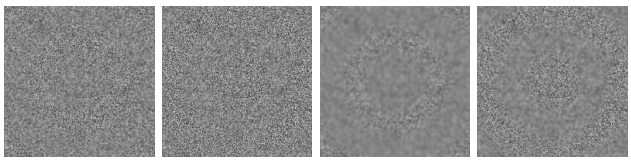

Figure 2. Case #1. First and second from the left: two synthetically generated images with $\gamma = 100$. Third and fourth from the left: two synthetically generated images with $\gamma = 300$.

- **sparse:** When we use less than 5 defocused images;
- **dense:** When we use more than 100 images.

In both cases each image has a size of $256pixels \times 256pixels$. Our results are obtained with volumes concentrated around a surface (depth map). Once the volume is estimated through the algorithm in Table 1, we calculate the radiance and the surface using (3.17) and (3.18).

Case #1: Dense Set of Defocused Images
We generate a *dense set* of 128 defocused images of a volume density concentrated around the surface shown in Figure 4 (top-right). The images are equally spaced between the two planes $1/v = 1/F - 1/.5m$ and $1/v = 1/F - 1/.9m$. In Figure 2 we show examples of such images for two different values of the pixel size ($\gamma = 100pixel/m$ and $\gamma = 300pixel/m$). Notice that the pixel size affects how information stored in the volume density is then encoded in the defocused images. An example of this effect can be seen in Figure 3, where we show 3 plane sections of the original volume and the relative estimates given by the algorithm. The one obtained with a "large" γ yields a better volume estimate than the one relative to a "small" γ. In Figure 4 we show the original radiance/surface and the reconstructed one for different values of γ. Finally, in Figure 5 we show the dependence of the reconstruction error (norm of the difference of the true and reconstructed volume densities) on the Gaussian additive noise in the input images.

Case #2: Sparse Set of Defocused Images We generate a *sparse set* of 5 defocused images of a volume density concentrated around the surface shown in Figure 8 (top-right). As in the case #1 the images are equally spaced between the two planes $1/v = 1/F - 1/.5m$ and $1/v = 1/F - 1/.9m$. In Figure 6 we can see some images of the stack for two different values of the pixel size ($\gamma = 100pixel/m$ and $\gamma = 300pixel/m$). In Figure 7 we show 3 plane sections of the original volume and the relative estimates given by the algorithm. The one obtained with a "large" γ yields a

Step	Action
1	Compute (analytically, if possible) the Fourier transforms of the PSF ϕ. If the PSF is a Gaussian, then we readily obtain the unnormalized Gaussian $\hat{\phi}(\mathbf{f}, \bar{\sigma}_v) = e^{-\frac{\bar{\sigma}_v^2}{2}\|\mathbf{f}\|^2}$;
2	At each frequency $\mathbf{f}$, build the operator $\widehat{\boldsymbol{\phi}}(\mathbf{f}, \cdot, \cdot)$ defined in eq. (3.8) for N input images and at N_{x_3} depth levels;
3	At each frequency $\mathbf{f}$, compute the pseudo-inverse $\widehat{\boldsymbol{\phi}}^{\dagger}$ of $\widehat{\boldsymbol{\phi}}$;
4	Apply each pseudo-inverse $\widehat{\boldsymbol{\phi}}^{\dagger}$ to the N-dimensional input images vector $\hat{\mathbf{I}}$, defined in eq. (3.9), and call the result $\widehat{W^0}(\mathbf{f}, f_3)$;
5	If no additional constraints are available, then the volume density $W(\mathbf{x}, x_3)$ is the inverse Fourier transform of $\widehat{W^0}(\mathbf{f}, f_3)$; if additional constraints are available, then use them to obtain $\mathbf{a}$ and compute $\widehat{W}(\mathbf{f}, f_3)$ as in eq. (3.11).

Table 1. Volume density estimation algorithm.

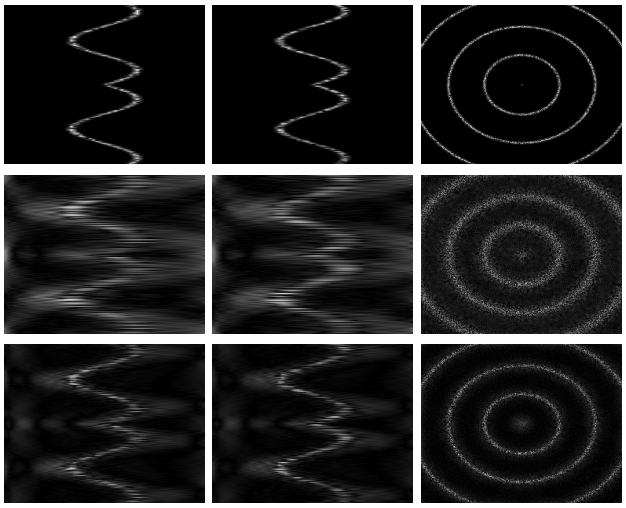

Figure 3. Case #1. Ground truth and reconstructed volume for 128 defocused images. Top row: section $x_1 - x_3$ of the ground truth volume density. Middle row: the same section from the reconstructed volume density with $\gamma = 100$. Bottom row: the same section from the reconstructed volume density with $\gamma = 300$.

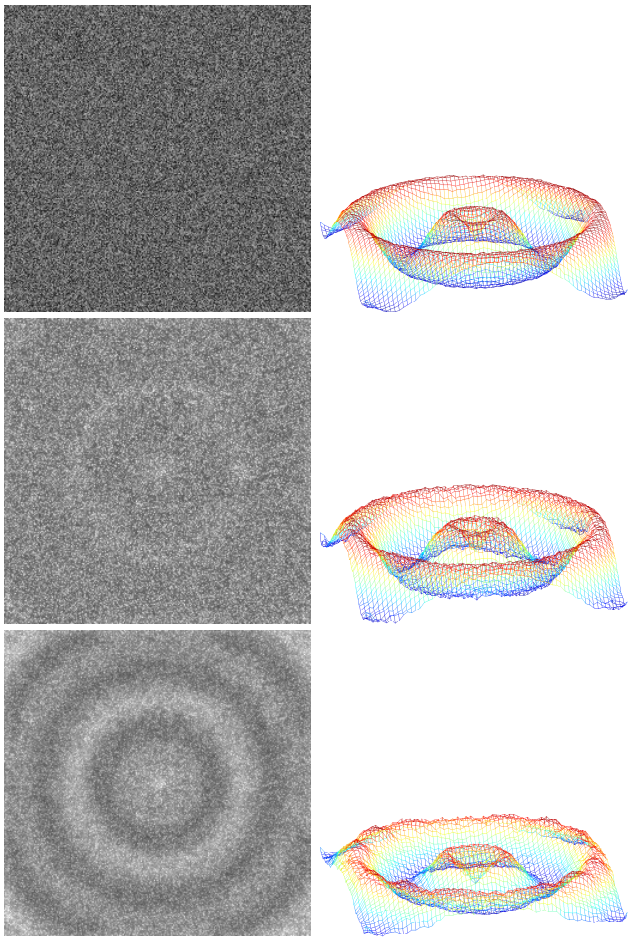

Figure 4. Case #1. Top row: Ground truth radiance (left) and relative depth map (right). Middle row: reconstructed radiance (left) and depth map (right) for $\gamma = 100$. Bottom row: reconstructed radiance (left) and depth map (right) for $\gamma = 300$.

better volume estimate than the one relative to a "small" γ. In Figure 8 we show the original radiance/surface and the reconstructed one for different values of γ. Finally, in Figure 9 we show the dependence of the reconstruction error (norm of the difference of the true and reconstructed volume densities) on the Gaussian additive noise in the input images.

5.2. Real Data

We also test the algorithm on the real data shown in Figure 10. The camera settings for the data sets are available at [13] and [5]. In Figure 10 on the first and third rows we show the two input image pairs. On the second and fourth rows we show the estimated depth map obtained by our algorithm (left) and the estimated depth map obtained by employing the least-squares algorithms in [5] (right). Notice

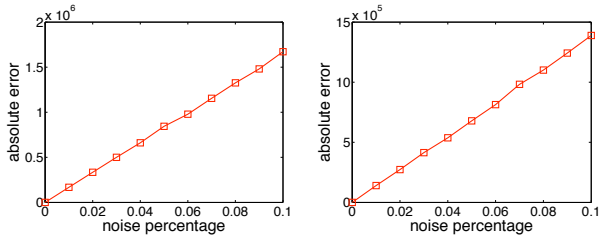

Figure 5. Case #1. Absolute reconstruction error for a dense stack of 128 defocused images corrupted by different noise levels. The graph on the left is relative to a $\gamma = 100$ while for the one on the right $\gamma = 300$.

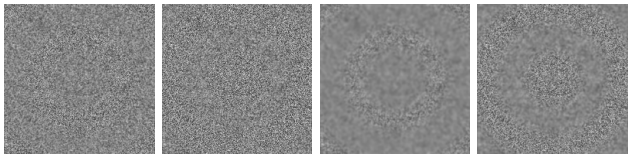

Figure 6. Case #2. First and second from the left: Two input images with $\gamma = 100$. Third and fourth from the left: Two input images with $\gamma = 300$.

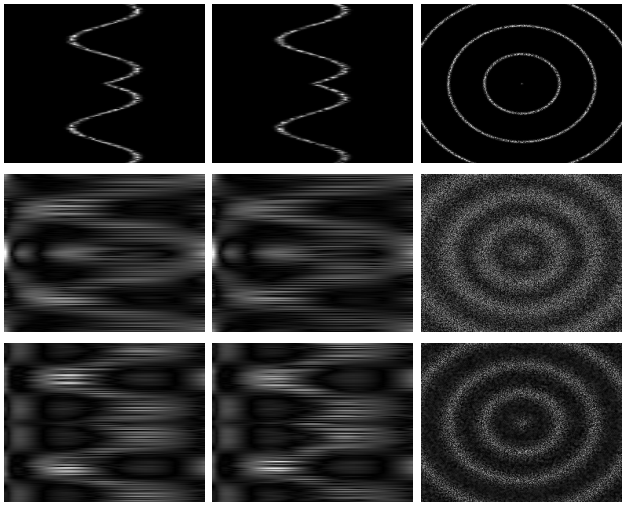

Figure 7. Case #2. Ground truth and reconstructed volume for 5 defocused images. Top row: section $x_1 - x_3$ of the ground truth volume density. Middle row: the same section from the reconstructed volume density with $\gamma = 100$. Bottom row: the same section from the reconstructed volume density with $\gamma = 300$.

that the surface estimates in the two cases are qualitatively similar, although the estimates obtained via the volumetric method tend to have a lower accuracy. This is due to the fact that while in the least-squares method the solution is constrained to be locally a plane, in our solution we simply use eq. (3.11) and the method summarized in Table 1, without imposing additional constraints on the solution. The advantage of our method however, is that it can be readily

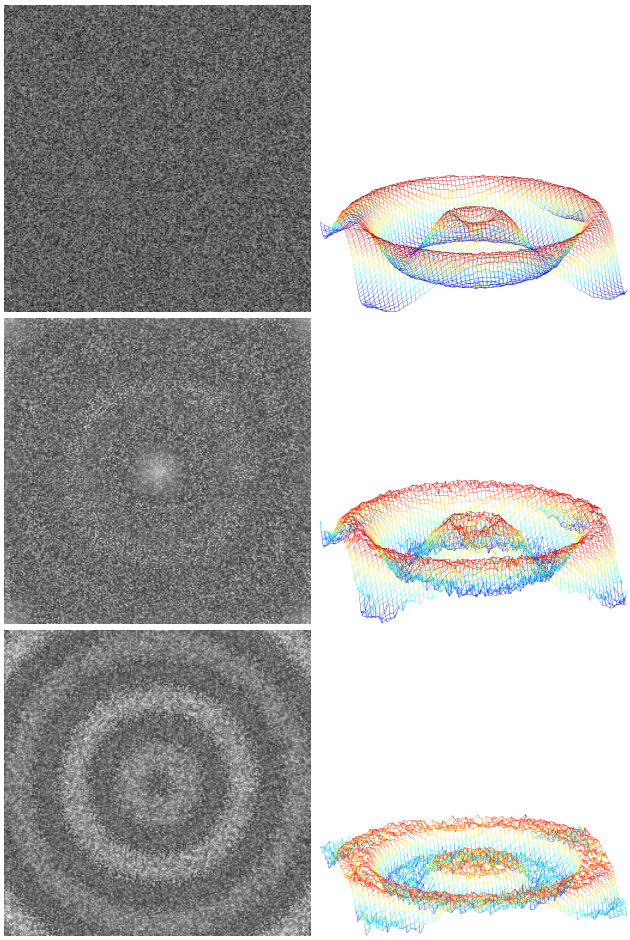

Figure 8. Case #2. Top row: Ground truth radiance (left) and relative depth map (right). Middle row: reconstructed radiance (left) and depth map (right) for $\gamma = 100$. Bottom row: reconstructed radiance (left) and depth map (right) for $\gamma = 300$.

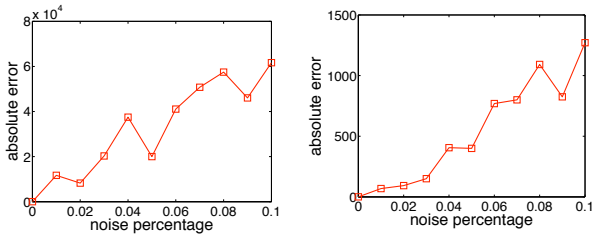

Figure 9. Case #2. Absolute reconstruction error for a stack of 5 defocused images corrupted by different noise levels. The graph on the left is relative to a $\gamma = 100$ while for the one on the right $\gamma = 300$.

applied to data containing multiple surfaces.

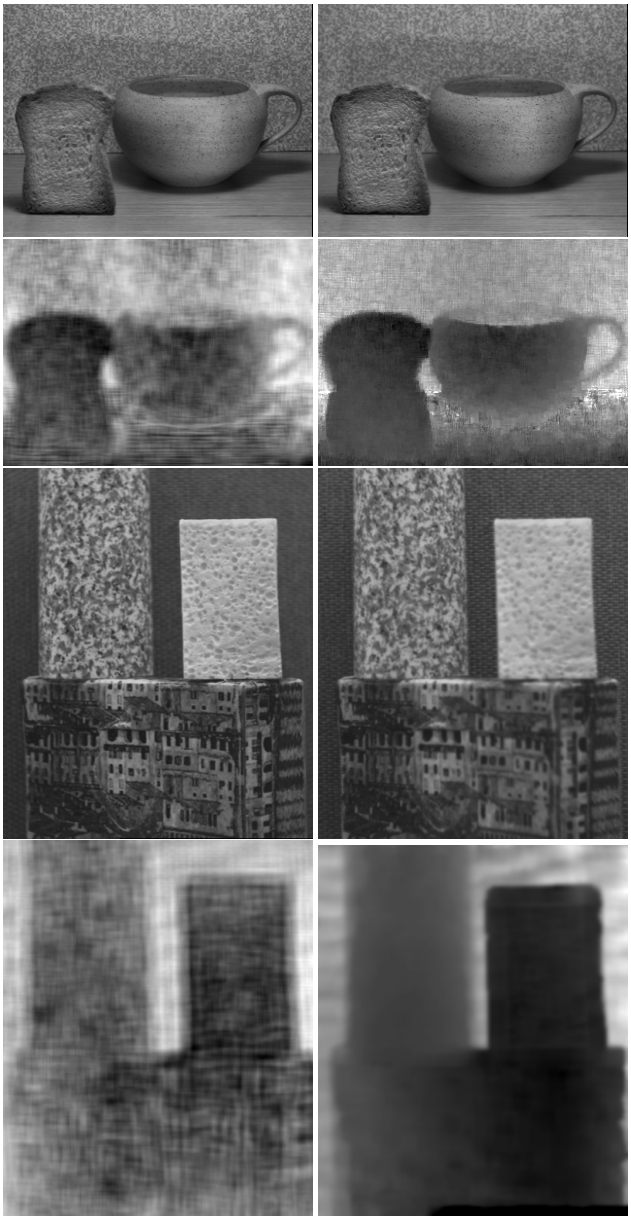

Figure 10. Real data set. First row: Two defocused images from [13]. Second row: Reconstructed depth map by using the proposed method (left) and the least-squares method described in [5] (right). Third row: Two defocused images from [5]. Fourth row: Reconstructed depth map by using the proposed method (left) and the least-squares method described in [5] (right).

6. Conclusions

We have presented novel analysis of the problem of volume density estimation from a stack of defocused images. We consider two cases: one where a dense set of images is available, and one where a sparse set of defocused images is available. In the first case we reduce the imaging model to a 3-D convolution, which allows to predict the maximum accuracy achievable both in space and frequency. In the second case we propose an efficient solution based on the singular value decomposition in the frequency domain. We explicitly characterize the space of all solutions and show how this can be constrained if prior knowledge about the scene is available. Experiments on both real and synthetic data validate the theory.

Acknowledgements

Alessandro Duci was supported by EU 6th Framework Program Grant MRTN-CT-2005-019481 ("From FLIM to FLIN" Marie-Curie Research Training Network) and Paolo Favaro was supported by EPSRC - EP/F023073/1(P).

References

[1] D. Agard. Optical sectioning microscopy: cellular architecture in three dimensions. *Annual Reviews of Biophysics and Bioengineering*, 13:191–219, 1984.

[2] M. Bertero and P. Boccacci. *Introduction to inverse problems in imaging.* Institute of Physics Publications, 1998.

[3] S. Chaudhuri and A. Rajagopalan. *Depth from defocus: a real aperture imaging approach.* Springer Verlag, 1999.

[4] T. Corle and G. Kino. *Confocal Scanning Optical Microscopy and Related Imaging Systems*. Academic Press, 1996.

[5] P. Favaro and S. Soatto. *3-d shape reconstruction and image restoration: exploiting defocus and motion-blur.* Springer-Verlag, 2006.

[6] S. W. Hasinoff and K. N. Kutulakos. Confocal stereo. *European Conf. on Computer Vision*, pages 620–634, 2006.

[7] A. Kak and M. Slaney. *Principles of Computerized Tomographic Imaging*. IEEE Press, 1988.

[8] M. Levoy, R. Ng, A. Adams, M. Footer, and M. Horowitz. Light field microscopy. In *ACM SIGGRAPH*, pages 924–934, New York, NY, USA, 2006. ACM Press.

[9] A. P. Pentland. A new sense for depth of field. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(4):523–531, July 1987.

[10] Y. Schechner and N. Kiryati. The optimal axial interval in estimating depth from defocus. *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, 2:843–848 vol.2, 1999.

[11] D. Snyder, T. Schulz, and J. O'Sullivan. Deblurring subject to nonnegativity constraints. *Signal Processing, IEEE Transactions on [see also Acoustics, Speech, and Signal Processing, IEEE Transactions on]*, 40(5):1143–1150, May 1992.

[12] M. Subbarao. Parallel depth recovery by changing camera parameters. In *Proc. IEEE Int. Conf. on Computer Vision*, page 149:55, Florida, USA, 1988.

[13] M. Watanabe and S. K. Nayar. Rational filters for passive depth from defocus. *Int. J. Comput. Vision*, 27(3):203–225, 1998.