Off-Axis Aperture Camera: 3D Shape Reconstruction and Image Restoration

Qingxu Dou
Heriot-Watt University
Edinburgh, UK
qd5@hw.ac.uk
http://www.eps.hw.ac.uk/~qd5

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

Abstract

In this paper we present a novel 3D surface and image reconstruction method based on the off-axis aperture camera. The key idea is to change the size or the 3-D location of the aperture of the camera lens so as to extract selected portions of the light field of the scene. We show that this results in an imaging device that blends defocus and stereo information, and present an image formation model that simultaneously captures both phenomena. As this model involves a non trivial deformation of the scene space, we also introduce the concept of scene space rectification and how this helps the reconstruction problem. Finally, we formulate our shape and image reconstruction problem as an energy minimization, and use a gradient flow algorithm to find the solution. Results on both real and synthetic data are shown.

1. Introduction

Reconstructing the 3-D geometry of a scene is one of the fundamental problems in Computer Vision and it has been studied for the past two decades [2]. This problem has been investigated mostly at scales that range between 1 and 100 meters [10]. At smaller scales, however, additional distortions, including out-of-focus blur, become dominant. Thus, conventional camera models, such as the pinhole model, are no longer suitable to solve a 3-D reconstruction task. In these scenarios, imaging models that take into account the finite aperture of the lens have been developed, as well as methods to recover both geometry and radiometry of the scene, such as shape from focus and defocus [1, 3, 4]. It is important to bear in mind that addressing 3-D reconstruction at smaller scales is not just a mere scientific exercise, but a fundamental problem in several applications, such as 3-D endoscopy and 3-D microscopy [11, 6]. More importantly, in most of these applications camera motion is constrained by the surrounding environment, such as in en-

doscopy, and, therefore, even if one wanted to apply stereo reconstruction methods, it would simply be not possible to employ a stereo camera or move the camera sideways.

An alternative to stereo that is applicable in constrained environments is shape from defocus, but the method tends to be more sensitive to the resolution of the image intensities. Our work is driven by the advantages of both methods. We propose to employ the *off-axis aperture (OAA) camera*, where several images are captured by moving an aperture in front of the camera lens as illustrated in Figure 1. By changing the aperture location we obtain images as if we were using a camera with a small lens and displaced with respect to the larger lens. This yields an effect similar to a stereo system. At the same time, due to the scale at play, defocusing effects are also visible. Therefore, in this paper we propose a novel image formation model that simultaneously captures both of these effects in one go (see section 2).

In the next section, we will see that defocus not only smoothes the image intensities, but also warps the projective geometry in a non-straightforward manner (see section 2.1). To facilitate the convergence of our reconstruction algorithms, we suggest to un-warp both the 3-D space and the image coordinates (section 3). This allows us to recover the geometry and the radiometry of the scene, by employing an efficient gradient flow algorithm (section 4).

1.1. Related Work

Off-axis apertures are not a novelty. In several fields, such as ophtalmology, astronomy, and microscopy, this kind of technique has been used to design imaging devices (*e.g.*, the stereo fundus camera [7, 14]). In astronomy, for instance, multi hole off-axis apertures have been widely used in testing optical elements for nearly one hundred years. A mask with two off-axis apertures, also called Hartmann Mask or Scheiner Disk, has been invented to correct of quantify defocus by the Jesuit astronomer Christoph Scheiner (1572-1650). This technique is based on the principle that, unless the scene is brought into focus, the two apertures will generate an effect similar to double vision.
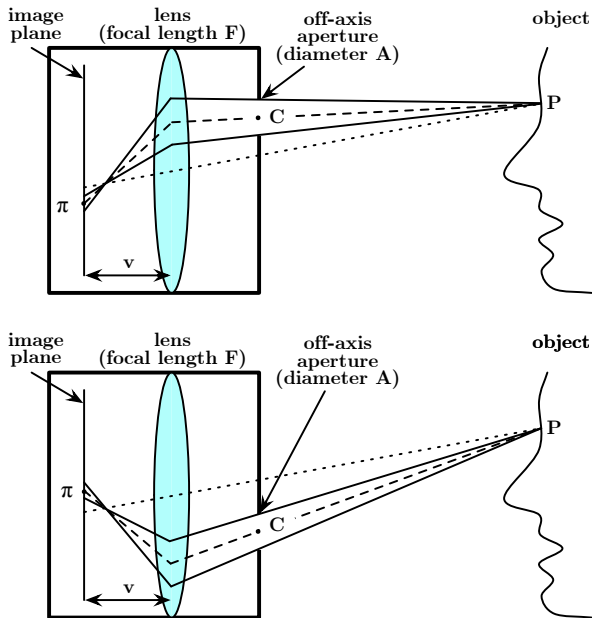
Figure 1. Geometry of the off-axis aperture camera. The off-axis aperture camera can be decomposed into three elements: An image plane, a lens, and the moving aperture. Top: Imaging a point P on a surface with the aperture positioned at the top. Bottom: Imaging the same point P as above with the aperture positioned at the bottom. Notice that when the point being imaged is not in focus, the image is a blurred disc and its center changes with the lens parameters as well as the aperture.

Off-axis aperture imaging can also be seen as a case of coded-aperture imaging [15]. In this research area an aperture is designed to solve a certain task, such as allowing for deblurring from one image or depth estimation [5, 12]. This work falls also into the category of *programmable imaging* methods, and therefore relates to work of Zomet et al. [8]. However, we make use of a lens in our system and therefore our image formation model is very different from that of [8]. Simoncelli and Farid [9] use the aperture to obtain range estimates from multiple images of the scene. Their method however is based on several assumptions that we do not make in this paper: first, the aperture is not constrained to lie on the lens as in [9]; second, we are not restricted to apertures (masks) that yield an estimate of the derivative of the input images with respect to the viewing position; third, we do not make the local fronto-parallel assumption.

2. Off-Axis Aperture Imaging

In this section we introduce the off-axis aperture camera and how this device can be used to capture 3D information about the scene. The main advantage of this camera is that no lens or image plane movement is required, as well as no

relative motion between camera and object is needed. We obtain images by changing the position and diameter of the aperture. While this can be achieved by physically moving a mask in front of the lens (as we do in our experiments), one could also place an LCD transparent display and turn its pixels on or off in space and in time so as to form the desired attenuation mask, as it has been done in [16].

2.1. Image Formation Model

In Figure 1 we sketch the off-axis aperture camera as a device composed of an image plane, a thin lens of focal length F,[1] and a moving aperture with center $\mathbf{C} = [C_1 \ C_2 \ C_3]^T \in \mathbf{R}^3$ and diameter A. The distance image plane to lens is denoted by v. Let $\mathbf{P} = [P_1 \ P_2 \ P_3]^T \in \mathbf{R}^3$ be a point in space lying on the object; then, the projection $\pi : \mathbf{R}^3 \mapsto \mathbf{R}^2$ of $\mathbf{P}$ on the image plane is defined by

$$\pi[\mathbf{P}] \doteq -v \frac{P_{1,2}}{P_3} + \left(1 - \frac{v}{v_0}\right) \frac{C_{1,2}P_3 - C_3 P_{1,2}}{P_3 - C_3} \quad (1)$$

where $v_0 = \frac{FP_3}{P_3 - F}$. Notice that when the image plane is at a distance $v = v_0$ or when the aperture is centered with respect to the lens, *i.e.*, $\mathbf{C} = [0 \ 0 \ 0]^T$, then the projection π coincides exactly with the perspective projection of $\mathbf{P}$ in a pinhole camera. Otherwise, as we move the aperture center $\mathbf{C}$, the projection $\pi[\mathbf{P}]$ is a *shifted* perspective projection. Furthermore, when $v \neq v_0$, the point $\mathbf{P}$ generates a blur disc of diameter B. By computing the projection of P through the boundary of the off-axis aperture, we find that the blur diameter B is

$$B \doteq A \left|1 - \frac{v}{v_0}\right| \frac{P_3}{P_3 - C_3} \quad (2)$$

which is identical to the well-known formula used in shape from defocus when $C_3 = 0$, *i.e.*, when the aperture lies on the lens [1, 3]. To ease the reading of the paper we summarize in Table 1 the symbols introduced so far.

Now, if we approximate the blur disc with a Gaussian (but a Pillbox function would do as well), we can explicitly write the intensity I measured on the surface of a square pixel $\mathbf{y}_{k,l} \in \mathbf{Z}^2$ on the image plane, and denoted by $\square_{k,l}$, as

$$I(\mathbf{y}_{k,l}) = \int_{\square_{k,l}} \int_{\mathbf{R}^2} \frac{e^{-\frac{\|\bar{\mathbf{y}} - \gamma^{-1}\pi[\mathbf{P}(\mathbf{x})]\|^2}{2\sigma^2(\mathbf{x})}}}{2\pi\sigma^2(\mathbf{x})} r(\mathbf{x}) d\mathbf{x} d\bar{\mathbf{y}} \quad (3)$$

where σ is related to the blur diameter B via $\sigma \doteq \gamma^{-1}\kappa B$, γ is the length of a side of a square pixel and κ is a calibration parameter.[2] $r(\mathbf{x})$ is the radiance emitted from a point

---

[1]The thin lens satisfies the following conjugation law $\frac{1}{v_0} + \frac{1}{P_3} = \frac{1}{F}$ where v_0 is the distance image plane to lens such that the point at depth P_3 is in focus.

[2]In our experiments $\kappa = 1/6$.

Table 1. Symbols used in the camera model and their meaning.

F	focal length
C	3-D position of the aperture center
v	distance image plane to lens
v_0	distance image plane to lens when object is in focus
P	3-D position of a point in space lying on the object
π	projection from the 3-D space to the image plane
h	point spread function (PSF)
B	blur diameter (referring to Pillbox PSF)
σ	spread of the blur disc (generic PSF)
A	aperture diameter
γ	size of the side of a pixel in mm
κ	calibration parameter
r	object radiance
I	measured image

$P(\mathbf{x})$ on the surface. We parametrized the surface of the object as $P : \mathbf{R}^2 \mapsto \mathbf{R}^3$. For now we will not make such parametrization explicit as it will be thoroughly analyzed in section 3. Rather, we make our representation of the radiance explicit via the coefficients $r_{i,j}$:

$$r(\mathbf{x}) \doteq \sum_{i,j} r_{i,j} U(\mathbf{x} - \mathbf{x}_{i,j}) \qquad (4)$$

where $U : \mathbf{R}^2 \mapsto \{0, 1\}$ denotes the indicator function, *i.e.*, $U(\mathbf{x})$ is 1 if and only if $-0.5 \leq x_1 < 0.5$ and $-0.5 \leq x_2 < 0.5$. The coordinates $\mathbf{x}_{i,j}$ belong to a regular lattice with step 1 in $\mathbf{Z}^2$. By substituting eq. (4) in eq. (3), we obtain

$$
\begin{aligned}
I(\mathbf{y}_{k,l}) &= \sum_{i,j} r_{i,j} \int_{\square_{k,l}} \int_{\square_{i,j}} h(\mathbf{x}, \bar{\mathbf{y}}) d\mathbf{x} d\bar{\mathbf{y}} \\
&= \sum_{i,j} r_{i,j} H_{i,j}(\mathbf{y}_{k,l})
\end{aligned}
\qquad (5)
$$

where $h(\mathbf{x}, \bar{\mathbf{y}}) \doteq \frac{1}{2\pi\sigma^2(\mathbf{x})} e^{-\frac{\|\bar{\mathbf{y}} - \gamma^{-1}\pi[\mathbf{P}(\mathbf{x})]\|^2}{2\sigma^2(\mathbf{x})}}$ denotes the *point spread function* of the camera, and $H_{i,j}(\mathbf{y}_{k,l})$ is implicitly defined by the above equation.

3. Space Warping: How to Bend Geometry to our Advantage

The image formation model in eq. (3) could be immediately used to reconstruct the 3-D shape of an object in the scene and its radiance. For instance, one could displace the center $\mathbf{C}$ of the aperture on a plane parallel to the image plane and apply standard stereo methods. However, when we only change C_3 or the aperture diameter A, the estimation problem becomes more difficult. The main issue is that the relationship between the input images captured with these modalities is highly non linear. To counteract such nonlinearities we suggest to change the representation of the unknowns so that their projection is as linear as possible. We call this method *warping*.

To illustrate the issues created by the original image formation model, we need to make our parametrization of the surface $\mathbf{P}(\mathbf{x})$ explicit. Suppose that $\mathbf{P}(\mathbf{x}) = [\mathbf{x}^T \; u(\mathbf{x})]^T$ where $u : \mathbf{R}^2 \mapsto [0, \infty)$ is the depth map of the scene. Then, a small variation δu of the depth map u causes the projection π to vary of

$$
\begin{aligned}
\delta\pi[\mathbf{P}(\mathbf{x})] &\doteq \delta u(\mathbf{x})\pi'[\mathbf{P}(\mathbf{x})] \\
&= \delta u(\mathbf{x})\frac{(\mathbf{x} - C_{1,2})(vF + C_3F - vC_3)}{F(u(\mathbf{x}) - C_3)^2}.
\end{aligned}
\qquad (6)
$$

One can observe that the variation $\delta\pi[\mathbf{P}(\mathbf{x})]$ in the above equation depends on the coordinates $\mathbf{x}$. In particular, as $\|\mathbf{x} - C_{1,2}\|$ grows, also $\|\delta\pi[\mathbf{P}(\mathbf{x})]\|$ grows. While in principle this behavior is acceptable, in practice it is an issue when using gradient-based techniques; part of the gradient of the cost functional will depend on $\delta\pi[\mathbf{P}(\mathbf{x})]$ and, as a consequence, the convergence to the solution will be unstable away from the center of the aperture (see Figure 8).

We suggest a simple method to offset this behavior. The key idea is to choose a parametrization of the surface of the object $\mathbf{P}(\mathbf{x})$ such that the projection $\pi[\mathbf{P}(\mathbf{x})] = \alpha\mathbf{x}$ for some scalar $\alpha \neq 0$, and $\mathbf{P}(\mathbf{x})$ does not depend on the varying camera parameters. If such a projection exists, then the captured images can be easily warped into each other so that the only difference between them is the relative amount of defocus. Once the (warped) depth map u is recovered from the warped images, one has to undo the warping by using the parametrization $\mathbf{P}(\mathbf{x})$ (unwarping). For instance, in shape from defocus $\mathbf{C} = [0 \; 0 \; 0]^T$ and v changes between the input images. Because one has that $\delta\pi[\mathbf{P}(\mathbf{x})] = \delta u(\mathbf{x})\frac{v\mathbf{x}}{u^2(\mathbf{x})}$, which is still dependent on $\mathbf{x}$, we cannot use the parametrization $\mathbf{P}(\mathbf{x}) = [\mathbf{x}^T \; u(\mathbf{x})]^T$. Let us define $\mathbf{P}(\mathbf{x}) \doteq [\mathbf{x}^T \; 1]^T u(\mathbf{x})$. Then, the projection $\pi[\mathbf{P}(\mathbf{x})] = -v\mathbf{x}$ and $\delta\pi[\mathbf{P}(\mathbf{x})] = 0$; the input images can be mapped into each other by warping their image domain (*i.e.*, scaling each image by its corresponding v and γ)

$$
\begin{aligned}
\hat{I}(\mathbf{z}_{k,l}, v) &\doteq \gamma^{-2}v^2 I(\gamma^{-1}v\mathbf{z}_{k,l}) \\
&= \gamma^{-2}v^2 \sum_{i,j} r_{i,j} \int_{\square_{k,l}} \int_{\square_{i,j}} \frac{e^{-\frac{\|\gamma^{-1}v\bar{\mathbf{y}} - \gamma^{-1}v\mathbf{x}\|^2}{2\sigma^2(\mathbf{x})}}}{2\pi\sigma^2(\mathbf{x})} d\mathbf{x} d\bar{\mathbf{y}} \\
&= \sum_{i,j} r_{i,j} \int_{\square_{k,l}} \int_{\square_{i,j}} \frac{1}{2\pi\hat{\sigma}^2(\mathbf{x})} e^{-\frac{\|\bar{\mathbf{y}} - \mathbf{x}\|^2}{2\hat{\sigma}^2(\mathbf{x})}} d\mathbf{x} d\bar{\mathbf{y}}
\end{aligned}
\qquad (7)
$$

Finally, once $u(\mathbf{x})$ is reconstructed, one needs to apply the parametrization $[\mathbf{x}^T \; 1]u(\mathbf{x})$ to obtain the correct depth map.

Remark 1 *Most algorithms for shape from defocus use implicitly this warping of the depth map. In fact, typically one*

works with images that have been aligned (warped) and the projection $\pi[\mathbf{P}(\mathbf{x})]$ is always approximated by $\mathbf{x}$. However, in most methods the last step (unwarping) is not applied thus preventing the algorithms from reconstructing the correct object. It is interesting to note that in the case of telecentric optics [13] some further simplifications are possible. We have that $\mathbf{C} = [0\ 0\ F]$, and hence $\pi[\mathbf{P}] \doteq -\frac{F\mathbf{x}}{u(\mathbf{x})-F}$ does not change between the images as it does not depend on v. With telecentric optics a warping between the input images is not required (as the projections do not change with v). However, as the projection is still a function of the depth map u and $\mathbf{x}$, the parametrization still needs to be changed (e.g., by setting $\mathbf{P}(\mathbf{x}) \doteq [\mathbf{x}^T\ 1]^T(u(\mathbf{x}) - F)$) and the corresponding reconstruction still needs to be unwarped.

In the case of varying A, the analysis is fairly straightforward as the projection π does not depend on A. Hence, we can choose $\mathbf{P} \doteq [P_{1,2}\ P_3]^T$ where

$$\mathbf{P}_{1,2}(\mathbf{x}) \doteq \frac{F(u(\mathbf{x}) - C_3)\mathbf{x} - (Fu(\mathbf{x}) - vu(\mathbf{x}) + vF)C_{1,2}}{C_3 v - C_3 F - vF} \tag{8}$$

and $P_3 \doteq u(\mathbf{x})$. It is immediate to see that for such a parametrization the projection $\pi[\mathbf{P}(\mathbf{x})] = \mathbf{x}$. As in the case of telecentric optics, no warping of the input images is required.

The case of varying C_3 is instead more difficult. We approximate the distance of the aperture from the lens C_3 with an average value $\bar{C}_3$. Then, let $\mathbf{P}(\mathbf{x}) = [\mathbf{x}^T\ \frac{u(\mathbf{x})}{u(\mathbf{x})-\bar{C}_3}]^T(u(\mathbf{x}) - \bar{C}_3)$ so that

$$\pi[\mathbf{P}(\mathbf{x})] \simeq -\frac{vF + (F - v)C_3}{F}\mathbf{x} + \left(1 - \frac{v}{v_0(\mathbf{x})}\right)\frac{u(\mathbf{x})C_{1,2}}{u(\mathbf{x}) - C_3}. \tag{9}$$

The variation $\delta\pi[\mathbf{P}(\mathbf{x})] \simeq \frac{\delta u(\mathbf{x})C_{1,2}}{F(u(\mathbf{x})-C_3)^2}(C_3(v - F) - vF)$ is then approximately independent of $\mathbf{x}$ (although it remains dependent on $u(\mathbf{x})$) and the input images can be pseudo-aligned simply by scaling the image domain by $\frac{vF+(F-C_3)}{F}$. After the pseudo-alignment, the resulting projection is then a function

$$\pi[\mathbf{P}(\mathbf{x})] \simeq -\mathbf{x} + \frac{1 - u(\mathbf{x})\left(\frac{1}{F} - \frac{1}{v}\right)}{1 - C_3\left(\frac{1}{F} - \frac{1}{v}\right)}\frac{C_{1,2}}{u(\mathbf{x}) - C_3}. \tag{10}$$

Remark 2 *Notice that in general it is not straightforward to find a parametrization $\mathbf{P}$ so that the projection does not depend on the depth map u. In such case we can approximately offset some of the undesired behavior due to the variation of the projection with respect to variations of the depth map.*

4. A Gradient Flow Algorithm

To find the unknown depth map u and the radiance r, we pose the following minimization problem

$$\begin{aligned}\hat{u}, \hat{r} \ = \ &\arg\min_{u,r} \sum_{n=1}^{N} \sum_{k,l} (I_n(\mathbf{y}_{k,l}) - J_n(\mathbf{y}_{k,l}))^2 \\ &+\lambda_1 \int \|\nabla u(\mathbf{x})\|^2 d\mathbf{x} \\ &+\lambda_2 \int \|r(\mathbf{x}) - r^*(\mathbf{x})\|^2 d\mathbf{x}\end{aligned} \tag{11}$$

where r^* is a reference radiance (*e.g.*, one of the input images or an average of the input images), and λ_1, λ_2 are positive scalars that control the amount of regularization. While the first term on the left hand side of eq. (11) matches the image model I_n to the measured images J_n for varying aperture centers C^n and diameters A^n, the second term introduces a smoothness constraint on the recovered depth map, and the third term prevents the estimated radiance from growing unboundedly.

The minimization is carried out by the following gradient flow algorithm[3]

$$\frac{\partial u(\mathbf{x}, t)}{\partial t} = -\nabla_u E \qquad \frac{\partial r(\mathbf{x}, t)}{\partial t} = -\nabla_r E \tag{12}$$

where the derivative with respect to iteration time is approximated by a forward difference, and $\nabla_u E$ and $\nabla_r E$ denote the gradient of the energy E with respect to the depth map u and the radiance r. The final computations of the gradients yields

$$\begin{aligned}\nabla_u E(\mathbf{x}_{i,j}) \ = \ &2 \sum_{n=1}^{N} \sum_{k,l} \Delta I_n(\mathbf{y}_{k,l}) I'_n(\mathbf{y}_{k,l}, \mathbf{x}_{i,j}) \\ &-2\lambda_1 \triangle u(\mathbf{x}_{i,j}) \\ \nabla_r E(\mathbf{x}_{i,j}) \ = \ &2 \sum_{n=1}^{N} \sum_{k,l} \Delta I_n(\mathbf{y}_{k,l}) H^n_{i,j}(\mathbf{y}_{k,l}) \\ &+2\lambda_2 (r(\mathbf{x}_{i,j}) - r^*(\mathbf{x}_{i,j}))\end{aligned} \tag{13}$$

where $\Delta I_n(\mathbf{y}_{k,l}) \doteq (I_n(\mathbf{y}_{k,l}) - J_n(\mathbf{y}_{k,l}))$ and

$$I'_n(\mathbf{y}_{k,l}, \mathbf{x}_{i,j}) \doteq \int_{\square_{k,l}} \int_{\square_{i,j}} h_n(\mathbf{x}, \bar{\mathbf{y}}) h'_n(\mathbf{x}, \bar{\mathbf{y}}) r(\mathbf{x}) d\mathbf{x} d\bar{\mathbf{y}} \tag{14}$$

where

$$\begin{aligned}h'_n(\mathbf{x}, \bar{\mathbf{y}}) \ \doteq \ &\frac{(\bar{\mathbf{y}} - \gamma^{-1}\pi[\mathbf{P}(\mathbf{x})])}{\sigma^2(\mathbf{x})}\gamma^{-1}\pi'[\mathbf{P}(\mathbf{x})]) \\ &+\left(\frac{\|\bar{\mathbf{y}} - \gamma^{-1}\pi[\mathbf{P}(\mathbf{x})]\|^2}{\sigma^3(\mathbf{x})} - \frac{1}{\sigma(\mathbf{x})}\right)\sigma'(\mathbf{x})\end{aligned} \tag{15}$$

---

[3]The gradient flow can be easily modified into more efficient schemes, without affecting the location of the minima, by premultiplying the gradient of the energy by any time-varying positive definite operator.

and

$$\sigma'(\mathbf{x}) = \gamma^{-1} \kappa A \, \mathrm{sign} \left(1 - \frac{v}{v_0(\mathbf{x})} \right) \frac{vC_3 - vF - C_3 F}{F(u(\mathbf{x}) - C_3)^2} \tag{16}$$

Remark 3 *As can be seen in most formulas, this algorithm can be implemented very efficiently. One can tabulate most computations from analytic solutions of the integrals of Gaussians over finite domains (i.e., via the error function) and by exploiting the separability of the Gaussian function. Indeed our current implementation in C++ with a Pentium $2GHz$ takes about 1 second to compute the above gradients on an image of dimensions 200×200 pixels.*

5. Experiments

5.1. Synthetic Data

In this section, we test the proposed algorithm on four synthetically-generated shapes: An equifocal plane, a scene made of equifocal planes at different depths (Cube data set), a slope (Slope data set), and a wave (Wave data set). For each of these shapes we show the reconstruction results obtained by either changing the diameter A of the aperture (see Figures 2 and 3) or the distance C_3 between the aperture and the lens (see Figures 5 and 6). In the experiments in Figures 2 and 3, we use 4 input images, which are obtained by setting the aperture diameter A as $2mm$, $4mm$, $5mm$, and $6mm$ and with other parameters fixed to focal length $F = 30mm$, $\mathbf{C} = [0\ 0\ 20mm]^T$, and distance image plane to lens $v = 45mm$. In Figure 2 we show only the input images corresponding to $A = 2mm$ and $A = 6mm$ (first and second from the left). In the same figure we show the true radiance (third from the left) and the estimated radiance (fourth from the left). In Figure 3 we show the ground truth for the shapes of the Cube, Slope, and Wave data sets on the left column and the corresponding estimated shapes on the right column. In the experiments in Figures 5 and 6, 3 input images are obtained by setting the distance of the aperture to the lens C_3 as $3mm$, $8mm$, and $12mm$ and with other parameters fixed to focal length $F = 30mm$, $A = 5mm$, and distance image plane to lens $v = 45mm$. As in the previous experiment, Figure 5 has 2 of the input images on the first and second illustration from the left, followed by the ground truth and estimated radiance. Similarly, in Figure 6 we show the ground truth shapes and the corresponding estimates on the data sets Cube, Slope, and Wave. In all experiments (including the ones on real data) the shape is initialized as a plane parallel to the image plane and the radiance is initialized as one of the input images. We find the algorithm somewhat insensitive to the initialization. To test the robustness of the method, we use 6 levels of additive Gaussian noise with, 0%, 1%, 2%, 5%, 15% and 20% of the radiance intensity. In the case of changing
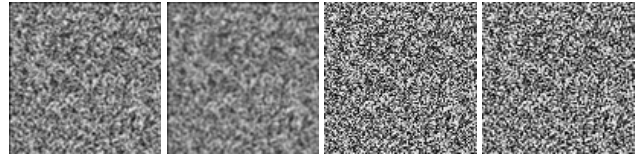

Figure 2. Image estimation with the change of the size of the aperture. First and second from the left: two of the input images in the case of the Cube dataset. Third and fourth from the left: the corresponding true radiance and the estimated radiance.

aperture diameter, in Figure 4 we show the absolute error (left), which we compute as the L_2 norm of the difference between the estimated shape and the true shape, and the relative error (right), which we compute as the absolute error and then normalize by dividing by the L_2 norm of the true shape. In the case of changing lens to aperture distance C_3, in Figure 7 we show the absolute error (left), which we compute as the L_2 norm of the difference between the estimated shape and the true shape, and the relative error (right), which we compute as the absolute error and then normalize by dividing by the L_2 norm of the true shape.

Warping. In section 3 we have introduced the concept of warping as a way to eliminate nonlinearities and therefore facilitate the convergence of the gradient-flow algorithm. To illustrate the advantage of warping the projection of the OAA camera we compare the surface estimated by using the original projection from the image formation model with the surface estimated by using the warped (rectified) projection. We consider the simplest instance of the original projections, i.e., the case of varying aperture A. In Figure 8 we show on the left the surface reconstructed when the original projection model is employed and the corresponding gradient flow is computed (see section 4); on the right we show the surface reconstructed when the warped projection model is employed in the gradient flow iteration. The improvement in the reconstruction is mostly due to the algebraic elimination of the projection nonlinearity which could introduce spurious local minima in the energy being minimized.

5.2. Real Data

To illustrate the algorithm above and validate it empirically, we test it on real images. Here we show an experiment of a scene containing a miniature house. The distance between the camera and the house model is about 400 millimeters. The images are captured by a Nikon D80 Digital Camera equipped with a Nikon AF Nikkor lens with $F = 50mm$, $v = 57.57mm$, and $\gamma = 23.7 pixels/mm$, by changing the aperture to $3mm$ and $22mm$. After correcting for the changes in intensity due to the difference in the aperture, we plot two of the input images on the top row of
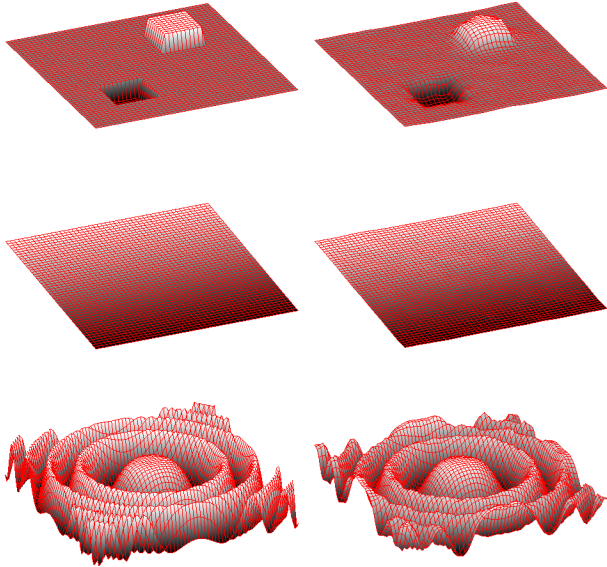
Figure 3. Shape reconstruction with the change of the size of the aperture. Left column (from top to bottom): the ground truth shape of the Cube, Slope and Wave date sets. Right column: The corresponding reconstructed shapes.
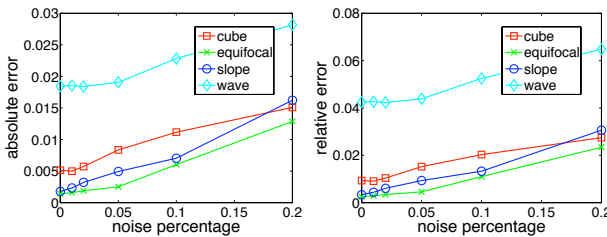

Figure 4. Error plots of the experiments with the change of the size of the aperture. Absolute and relative errors between the ground truths and the estimated depth maps of the four synthetic data sets for six noise levels.
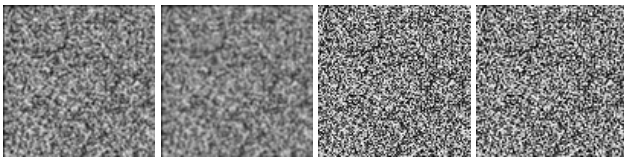

Figure 5. Image estimation by changing the distance of the aperture from the lens. First and second from the left: two of the input images in the case of the Cube dataset. Third and fourth from the left: the corresponding true radiance and the estimated radiance.

Figure 9. On the bottom row we show the restored radiance (left) and the recovered shape (right). The third row shows several views of the reconstructed scene. As one can see the qualitative shape has been successfully captured.
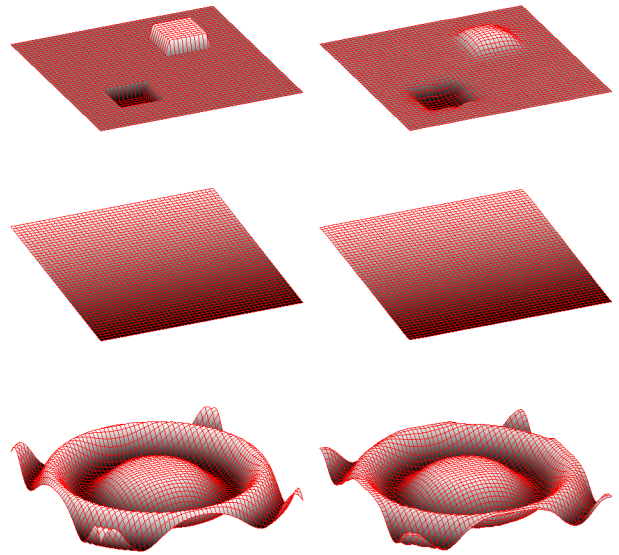

Figure 6. Shape reconstruction with the change of the distance from the aperture to the lens. Left column (from top to bottom): the ground truth shape of the Cube, Slope and Wave date sets. Right column: The corresponding reconstructed shapes.
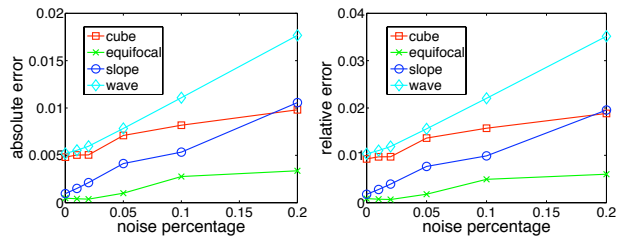

Figure 7. Error plots of the experiments with the change of the distance from the aperture to the lens. Absolute and relative errors between the ground truths and the estimated depth maps of the four synthetic data sets for six noise levels.
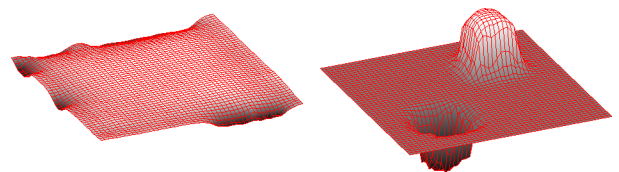

Figure 8. Comparison of the estimated shape when the original image formation model is employed (left) and when the projection is rectified via warping (right). The true shape is the same as the one shown at the top left of Figure 3.

6. Conclusions

We presented a novel family of shape estimation and image restoration algorithm based on the off-axis aperture
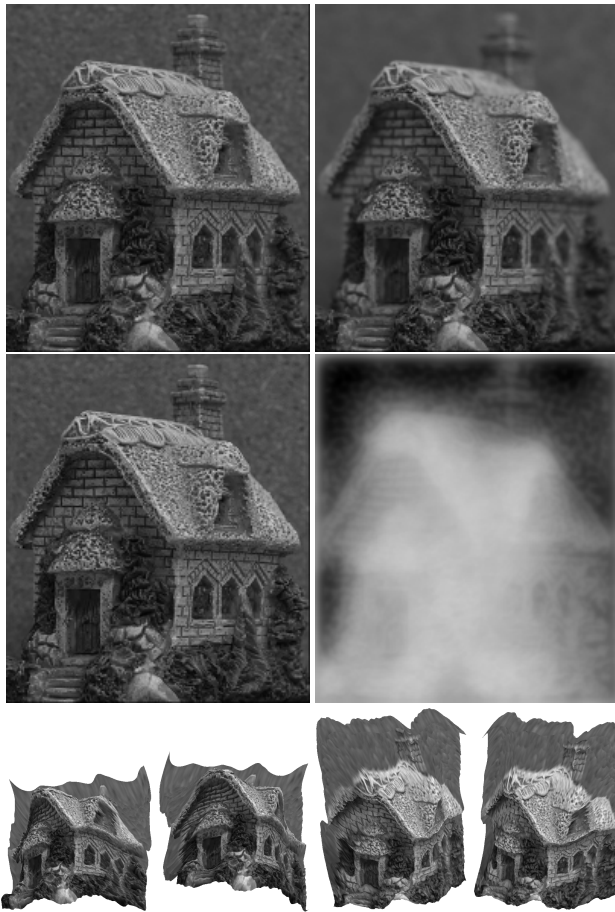
Figure 9. Experiments on real data. Top row: the two input images; the image on the left has been captured with $A = 3mm$, while the image on the right has been captured with $A = 22mm$. Second row: radiance (left) and estimated depth map in gray scale (right). Third row: several views of the reconstructed 3-D scene.

camera. The main property of this camera is that it does neither need a relative motion with respect to the object nor a change in the image plane and lens position. The only moving part is the aperture. We change its location and its diameter and show how this relates to the 3-D structure of the scene via the image formation model. We propose several parametrizations for the geometry of the object so that the resulting minimization algorithm is simple, and its convergence well-behaved. Results on several synthetic and real datasets are shown and demonstrate the effectiveness of the technique.

Acknowledgements

References

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

[2] O. Faugeras. *Three-dimensional computer vision: a geometric viewpoint.* MIT Press, Cambridge, MA, USA, 1993.

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

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

[5] A. Levin, R. Fergus, F. Durand, and W. T. Freeman. Image and depth from a conventional camera with a coded aperture. In *ACM SIGGRAPH '07*, page 70, New York, NY, USA, 2007.

[6] 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.

[7] T. Nanjo. Stereoscopic retinal camera including vertically symmetrical apertures. *Patent N.5302988*, April 1994.

[8] S. K. Nayar, V. Branzoi, and T. E. Boult. Programmable imaging using a digital micromirror array. *Computer Vision and Pattern Recognition*, 1(1063-6919):I–436–I–443 Vol.1, 2004.

[9] E. P. Simoncelli and H. Farid. Direct differential range estimation using optical masks. pages 82–93, 1996.

[10] N. Snavely, S. M. Seitz, and R. Szeliski. Photo tourism. *ACM Transactions on Graphics (SIGGRAPH Proceedings)*, 25(3):835–846, 2006.

[11] J. Strutz. 3d endoscopy. In *HNO*, volume 41, pages 128–130, 1993.

[12] A. Veeraraghavan, R. Raskar, A. Agrawal, A. Mohan, and J. Tumblin. Dappled photography: mask enhanced cameras for heterodyned light fields and coded aperture refocusing. In *ACM SIGGRAPH '07*, page 69, New York, NY, USA, 2007.

[13] M. Watanabe and S. K. Nayar. Telecentric optics for computational vision. pages 439–451, 1996.

[14] J. Xu, O. Chutatape, C. Zheng, and P. Kuan. Three dimensional optic disc visualization from stereo images via dual registration and ocular media optical correction. *British Journal of Ophthalmology*, 90:181–185, 2006.

[15] L. Zhang, B. K. P. Horn, and R. C. Lanza. Pattern design and imaging methods in 3-d coded aperture techniques. *Nuclear Science Symposium*, 3:1521–1524, 1998.

[16] A. Zomet and S. K. Nayar. Lensless imaging with a controllable aperture. *Computer Vision and Pattern Recognition*, 1(1063-6919):339–346, 2006.