Autocalibration and Uncalibrated Reconstruction of Shape from Defocus

Yifei Lou
UCLA
Los Angeles, USA
yflou@math.ucla.edu

Paolo Favaro
Heriot-Watt University
Edinburgh, UK
p.favaro@hw.ac.uk

Andrea L. Bertozzi
UCLA
Los Angeles, USA
bertozzi@math.ucla.edu

Stefano Soatto
UCLA
Los Angeles, USA
soatto@ucla.edu

Abstract

Most algorithms for reconstructing shape from defocus assume that the images are obtained with a camera that has been previously calibrated so that the aperture, focal plane, and focal length are known. In this manuscript we characterize the set of scenes that can be reconstructed from defocused images regardless of calibration parameters. In lack of knowledge about the camera or about the scene, reconstruction is possible only up to an equivalence class that is described analytically. When weak knowledge about the scene is available, however, we show how it can be exploited in order to auto-calibrate the imaging device. This includes imaging a slanted plane or generic assumptions on the restoration of the deblurred images.

1. Introduction

Images taken with a finite aperture, as opposed to a pinhole, exhibit spatial frequencies that depend on the shape of the scene, in addition to its radiance. It has been noticed long ago that, under suitable assumptions on the radiance, one could infer the depth map of a scene from a single image [12], or one could measure multiple images with different aperture settings and recover both the depth map and a "deblurred" image, i.e., the radiance of the scene [15]. Shape from defocus (SFD) has since become an active niche in computer vision, with relations to restoration and blind deconvolution in image processing, and applications to confocal microscopy, fine-scale metrology and high-definition photography [10]. A snapshot of the state of the art at various times has been captured by textbooks such as [3, 8] and a variety of recent algorithms have been proposed that concentrate on various aspects of the problem from accuracy (optimality of the reconstruction [14, 7]) to computational efficiency (multi-scale or real-time implementations [4, 17]) to the design of dedicated hardware [10].

In order to return an estimate of the shape of the scene (represented by its depth map) and of its radiance (the "deblurred" image), however, *all existing algorithms require knowledge of the internal parameters of the camera*, including its aperture, focal length, and position of the imaging plane. A variety of calibration procedures have been devised to measure such parameters, as well as to compensate for magnification and alignment artifacts due to the changes in imaging geometry caused by most common focusing mechanisms in commercial cameras. Most methods for camera calibration, for instance [16], are based on a pinhole model of the camera. However, camera calibration can also be obtained by exploiting defocus information, as in [1] and, more recently, in [2].[1]

It is well-known that, given images taken from pin-hole cameras at different vantage points, in the absence of calibration information we can reconstruct the geometry of the scene up to a global projective transformation [5]. It is also common practice to exploit partial knowledge of the scene, for instance the presence of parallel lines, to infer some or all of the calibration parameters, and to upgrade the reconstruction to knowledge up to an affine, or similarity transformation [9]. We also know that reconstruction up to the equivalence class induced by projective (or affine, or similarity) transformations is sufficient to accomplish certain tasks, such as hand-eye coordination or reprojection of the scene onto a novel vantage point. What about images taken from the same vantage point, but with different apertures or focal lengths? Can we *describe the equivalence class of scenes that are indistinguishable from defocused images* in the absence of calibration information? And can we use *generic knowledge about the scene in order to* reduce this equivalence class, and *upgrade the reconstruction to affine or Euclidean?* Finally, can reconstruction up to this equivalence class be useful for other tasks; for instance, *can we still restore an image* from an uncalibrated reconstruction?

In this paper we address these questions as follows: We give an analytical characterization of the reconstruction of shape from defocus in the absence of calibration data (*un-*

[1] Camera parameters are fundamental in shape from defocus; this is shown in [13] where the set of camera parameters that yield optimal reconstruction is determined (based on the Cramer-Rao bound of the variance of the error in the estimate of blur).

calibrated reconstruction, Claim 1). We then show that knowledge of the presence of *planes* in the scene allows partial recovery of calibration parameters and allows upgrading the reconstruction (*autocalibration*, Claim 2). Finally, we show that uncalibrated reconstruction and partial knowledge of the camera parameters is sufficient to deblur the images (Sect. 4).

1.1. Formalization of the problem

Consider an image obtained with a camera with point-spread function (PSF) $h : \mathbb{R}^2 \to \mathbb{R}^+$ that has some controllable degrees of freedom, for instance the distance between the plane of the lens and the sensor as in the focusing mechanism of common commercial cameras. Usually such a controllable degree of freedom is related to the scale of the PSF, called *blur radius*[2] $b \in \mathbb{R}^+$ via

$$b(x) = \frac{Dv}{2}\left|\frac{1}{F} - \frac{1}{v} - \frac{1}{s(x)}\right| \tag{1}$$

where D is the (fixed) lens diameter, F is the (fixed) focal length, $s : \Omega \subset \mathbb{R}^2 \to \mathbb{R}^+$ is the shape of the scene, represented as the graph of a function with domain on the plane of the sensor, and v is the controllable lens parameter (the distance between the sensor and the lens). We indicate the image formation process as

$$I(x|\nu) = \int h(x, y|b(y))r(y)dy \quad x \in \Omega \subset \mathbb{R}^2 \tag{2}$$

where $r : \mathbb{R}^2 \to \mathbb{R}^+$ is the radiance of the scene and $\nu = \{D, F, v\}$ are the calibration parameters embedded in b. The conditions under which one can reconstruct the shape $s(\cdot)$ given two or more images $I(x|\nu_1), I(x|\nu_2)$ are described in [6], and involve assumptions on the radiance r (trivially, a constant radiance does not allow the reconstruction of the correct shape). However, even under these conditions, reconstruction of the correct shape requires knowledge of the calibration parameters ν_1, ν_2 that are characteristics of the imaging device and are not easily measurable. It is common to assume that the two images are obtained by changing only one of the calibration parameters, via the focusing mechanism, for instance $\nu_1 = \{D, F, v_1\}$ and $\nu_2 = \{D, F, v_2\}$.

The first question we address pertains to the degree in which a scene can be reconstructed in the absence of calibration information. More specifically, we want to characterize all scenes that are indistinguishable from the given data, in the sense of yielding the same images under different calibration parameters. We concentrate on the dependency of the reconstruction on calibration parameters, and assume that the conditions on the radiance described in [6] are satisfied, which we express by saying that the radiance is *admissible*.

[2]This name is because for pill-box PSF this is the actual radius of the pill-box.

Problem 1 *Let a scene with shape s and radiance r generate two or more images $I_j(x|\nu_j)$ under calibration parameters ν_j, $j = 1, \dots, N \geq 2$. Characterize the scenes $\tilde{s}$ that yield the same images I_j for some set of calibration parameters $\tilde{\nu}_j$, for all admissible radiances r.*

We call such scenes $\tilde{s}$ *indistinguishable* from s, and we characterize them in Claim 1. Note that the condition that the scenes be indistinguishable for all admissible radiances reflects our interest in characterizing the *generic ambiguities* in uncalibrated reconstruction. Of course, for any given radiance r one could construct different scenes that, for *that radiance,* yield identical images. These are *pathological ambiguities* that have rather limited interest since they are specific to a particular radiance.

More interesting is to ascertain whether there are generic properties of (part of) the scene that can be enforced in order to reduce the generic ambiguities. The simplest conceivable is *linearity*, i.e., the presence of planes in the scene:

Problem 2 *Determine the subset of indistinguishable scenes that preserve linearity.*

This problem is addressed in Claim 2, where we show that knowledge of the presence of planes in the scene can be exploited to *upgrade* the reconstruction and partial *autocalibration*. In Claim 3 we also show that adding additional views does not help reducing the unknowns. Finally,

Problem 3 *Characterize the class of deblurred images that can be retrieved from blurred ones in the absence of calibration information.*

This is addressed in Section 4. Although these results are proven analytically, to test the relevance of the claims we validate them also experimentally on synthetic as well as real images captured with ground-truth calibration in Section 5.

2. Uncalibrated reconstruction

Before we address Problem 1 we must guarantee the uniqueness of the relationship between shape and blur radius expressed in eq. (1). We do so by considering radiances that are *sufficiently exciting*.

Definition 1 *Let I_j, $j = 1, 2$ be obtained by a scene s, r and $\tilde{I}_j$, $j = 1, 2$ be obtained by $\tilde{s}, r$; then, the radiance r is* sufficiently exciting *if $I_j = \tilde{I}_j$ if and only if $b_j = \tilde{b}_j$, $\forall\, x \in \Omega \subset \mathbb{R}^2$, $j = 1, 2$.*

This definition guarantees that scenes with the same shape would yield identical images only if they were subject to the same amount of defocus. We therefore move on to the solution of Problem 1, summarized in the following claim.

Claim 1 *The set of scenes $\tilde{s}$ that are indistinguishable from a given scene s that is non-equifocal ($s(x) \neq \text{const.}$) belong to the following set*

$$\left\{ \tilde{s} : \mathbb{R}^2 \to \mathbb{R}^+ \;\middle|\; \frac{1}{\tilde{s}(x)} = \alpha \frac{1}{s(x)} + \beta,\ \forall x \right\} \tag{3}$$

where α, β are constants.

Proof: By substituting eq. (1) into $b_j = \tilde{b}_j, j = 1, 2$ and by replacing the absolute value, we get

$$\frac{1}{\tilde{s}(x)} = \pm \frac{D v_j}{\tilde{D} \tilde{v}_j} \left(\frac{1}{s(x)} + \frac{1}{v_j} - \frac{1}{F} \right) + \frac{1}{\tilde{F}} - \frac{1}{\tilde{v}_j}, \quad j = 1, 2 \tag{4}$$

It follows from (4) that there are 4 possible sign and index combinations, i.e.,

$$\begin{cases} \frac{1}{\tilde{s}(x)} = +\frac{D v_1}{\tilde{D} \tilde{v}_1} \left(\frac{1}{s(x)} + \frac{1}{v_1} - \frac{1}{F} \right) + \frac{1}{\tilde{F}} - \frac{1}{\tilde{v}_1} \\ \frac{1}{\tilde{s}(x)} = +\frac{D v_2}{\tilde{D} \tilde{v}_2} \left(\frac{1}{s(x)} + \frac{1}{v_2} - \frac{1}{F} \right) + \frac{1}{\tilde{F}} - \frac{1}{\tilde{v}_2} \end{cases} \tag{5}$$

$$\begin{cases} \frac{1}{\tilde{s}(x)} = +\frac{D v_1}{\tilde{D} \tilde{v}_1} \left(\frac{1}{s(x)} + \frac{1}{v_1} - \frac{1}{F} \right) + \frac{1}{\tilde{F}} - \frac{1}{\tilde{v}_1} \\ \frac{1}{\tilde{s}(x)} = -\frac{D v_2}{\tilde{D} \tilde{v}_2} \left(\frac{1}{s(x)} + \frac{1}{v_2} - \frac{1}{F} \right) + \frac{1}{\tilde{F}} - \frac{1}{\tilde{v}_2} \end{cases} \tag{6}$$

$$\begin{cases} \frac{1}{\tilde{s}(x)} = -\frac{D v_1}{\tilde{D} \tilde{v}_1} \left(\frac{1}{s(x)} + \frac{1}{v_1} - \frac{1}{F} \right) + \frac{1}{\tilde{F}} - \frac{1}{\tilde{v}_1} \\ \frac{1}{\tilde{s}(x)} = +\frac{D v_2}{\tilde{D} \tilde{v}_2} \left(\frac{1}{s(x)} + \frac{1}{v_2} - \frac{1}{F} \right) + \frac{1}{\tilde{F}} - \frac{1}{\tilde{v}_2} \end{cases} \tag{7}$$

$$\begin{cases} \frac{1}{\tilde{s}(x)} = -\frac{D v_1}{\tilde{D} \tilde{v}_1} \left(\frac{1}{s(x)} + \frac{1}{v_1} - \frac{1}{F} \right) + \frac{1}{\tilde{F}} - \frac{1}{\tilde{v}_1} \\ \frac{1}{\tilde{s}(x)} = -\frac{D v_2}{\tilde{D} \tilde{v}_2} \left(\frac{1}{s(x)} + \frac{1}{v_2} - \frac{1}{F} \right) + \frac{1}{\tilde{F}} - \frac{1}{\tilde{v}_2} \end{cases} \tag{8}$$

By assuming that $s(x)$ and $\tilde{s}(x)$ are not constant eq. (5) yields

$$\frac{D v_1}{\tilde{D} \tilde{v}_1} = \frac{D v_2}{\tilde{D} \tilde{v}_2} \tag{9}$$

$$\frac{D v_1}{\tilde{D} \tilde{v}_1} \left(\frac{1}{v_1} - \frac{1}{F} \right) + \frac{1}{\tilde{F}} - \frac{1}{\tilde{v}_1} = \frac{D v_2}{\tilde{D} \tilde{v}_2} \left(\frac{1}{v_2} - \frac{1}{F} \right) + \frac{1}{\tilde{F}} - \frac{1}{\tilde{v}_2}. \tag{10}$$

We can deduce that eqs. (9,10) hold either if $D = \tilde{D}$ and $\frac{v_1}{\tilde{v}_1} = \frac{v_2}{\tilde{v}_2}$ or if $v_1 = v_2$ and $\tilde{v}_1 = \tilde{v}_2$. The latter is trivially eliminated since we assume that the two images in each set are obtained with different values of v.

Similarly we can obtain conditions on focal settings according to the last three cases eqs. (6-8), but since all parameters must be positive, these ambiguities can be discarded. So, in order to generate the same images I_1 and I_2, the following constraints on calibration parameters must be satisfied

$$D = \tilde{D} \quad \text{and} \quad \frac{v_1}{\tilde{v}_1} = \frac{v_2}{\tilde{v}_2}. \tag{11}$$

This leads to

$$\frac{1}{\tilde{s}(x)} = \frac{\alpha}{s(x)} + \beta,\ \forall x \tag{12}$$

with $\alpha = \frac{v_1}{\tilde{v}_1}$ and $\beta = \frac{1}{\tilde{F}} - \frac{\alpha}{F}$.

3. Partial autocalibration

In this section we explore the extent in which unique reconstruction is possible even in the absence of calibration information. In order to further reduce the indistinguishable set one can exploit partial knowledge about the scene, or partial knowledge about the imaging device. We explore these two options in order.

3.1. Partial scene knowledge

As we have seen in the previous section, all scenes that are indistinguishable given a set of images because of lack of knowledge of calibration parameters are related via an affine transformation of the inverse depth. Because such a transformation does not preserve certain properties of the scene, we can attempt to use prior information to reconstruct the *correct* shape of the scene despite lack of knowledge of calibration parameters, in a way analogous to reconstruction from vanishing points in structure from motion [11].

Claim 2 *If $s(x) = a^T x + b$ with $a \neq 0$, i.e., if the scene is a slanted plane, then the indistinguishable set is a plane if and only if $\beta = 0$.*

Proof: Suppose the indistinguishable set has the form $\tilde{s}(x) = c^T x + d$. It follows from (12) that for all $x \in \Omega$ we must have

$$a^T x + b = \alpha (c^T x + b) + \beta (a^T x + b)(c^T x + d) \tag{13}$$

Then the quadratic term of the above equation must be zero, that is, $\beta (a c^T) = 0$. Since $a \neq 0$, we must have either $\beta = 0$ or $c = 0$. If $c = 0$ then $\tilde{s}(x) = d$ is constant; this implies that $\alpha = 0$, which is not admissible (see Claim 1).

By substituting back into eq. (12) we get

$$a^T x + b = \alpha (c^T x + d) \tag{14}$$

which means that enforcing planarity in the uncalibrated reconstruction yields the correct reconstruction up to a single scale α (i.e., the reconstruction is a plane parallel to the original one).

This suggests a procedure for autocalibration: Assuming that the scene contains one or more planes, perform reconstruction using arbitrary parameters, then adjust them until the uncalibrated reconstruction is planar. One could automate this process by introducing a complexity measure for the uncalibrated reconstruction so as to enforce planarity in a minimum-description length framework, but this is well beyond the scope of this paper. In Section 5.2 we give experimental evidence that planarity is not preserved when $\beta \neq 0$.

3.2. Multiple views

In addition to having partial knowledge about the scene, we could have partial knowledge about the cameras, for instance by knowing that more than two views are captured by the same camera (e.g., same D and F, the only change being the scalar parameters v_j.) Unfortunately, the following claim shows that using more images does not help reducing the ambiguity in the reconstruction.

Claim 3 *Multiple views cannot eliminate the affine ambiguity; i.e., given $N \geq 3$ images $I_j, j = 1, \ldots, N$ with different parameters v_j, the reconstruction is known up to 2 parameters, as described in Claim 1.*

Proof: Consider three images, $I_j, j = 1, 2, 3$. Suppose we can reconstruct a scene s_{ij} from images I_i, I_j for $i \neq j$, thus getting three depth maps. We will prove that all the depth maps s_{ij} are identical, which means that additional images do not carry information.

Since the depth maps s_{23} and s_{13} yield the same image I_3, we get

$$\frac{Dv_3}{2}\left|\frac{1}{F}-\frac{1}{v_3}-\frac{1}{s_{23}}\right| = \frac{Dv_3}{2}\left|\frac{1}{F}-\frac{1}{v_3}-\frac{1}{s_{13}}\right| \quad (15)$$

Let us assume that $s_{23} \neq s_{13}$. Then, from the above equation it follows that

$$\frac{1}{s_{23}}+\frac{1}{s_{13}} = 2\left(\frac{1}{F}-\frac{1}{v_3}\right) \quad (16)$$

If all the depth maps are different, then similarly to eq. (16), we obtain

$$\frac{1}{s_{13}}+\frac{1}{s_{12}} = 2\left(\frac{1}{F}-\frac{1}{v_1}\right) \quad (17)$$

and

$$\frac{1}{s_{12}}+\frac{1}{s_{23}} = 2\left(\frac{1}{F}-\frac{1}{v_2}\right). \quad (18)$$

By solving with respect to s_{ij} in the three equations (16-18), we obtain that all s_{ij} are constant, that is in contradiction with the assumption of Claim 1.

In the case where two depth maps are the same, while a third one is different from the first two, we assume that $s_{23} \neq s_{13}$ and $s_{13} = s_{12}$. Then, by comparing eq. (16) and eq. (18) we obtain

$$\frac{1}{F}-\frac{1}{v_3} = \frac{1}{F}-\frac{1}{v_2} \quad (19)$$

This implies that $v_2 = v_3$, that contradicts our initial assumption of having different focal settings.

The case of $N > 3$ follows by induction. Assume that the reconstructions obtained from any $N-1$ images yields the same depth map, and then consider an additional new image I_N. For any $i \neq j$, with $i, j = 1, \ldots, N-1$, three images I_i, I_j and I_N yield three depth maps s_{ij}, s_{iN}, s_{jN}. As shown in the first part, these three depth maps are identical. Since i, j are arbitrary, we conclude that adding a new image will not add more constraints to the depth map reconstruction.

4. Calibration parameters in image restoration

As we have seen, once two or more images are given, we can compute an uncalibrated reconstruction $\tilde{s}(x|\alpha, \beta)$ such that $\frac{1}{\tilde{s}} = \frac{\alpha}{s} + \beta$ for some arbitrary α, β. This can be done with any SFD algorithm, for instance those described in [3]. Once we have a reconstruction, albeit a wrong one, we can use it to obtain a "deblurred" version of the original images, or an estimate of the radiance of the scene. It is then immediate to show that

Claim 4 *If we are given 2 defocused images, the true lens aperture D and the ratio between v_1 and v_2, then image restoration can be performed exactly even if the true shape is unknown.*

Proof: If we obtain the depth map $\tilde{s}$ with camera parameters $\tilde{D}$, $\tilde{v}_1$ and $\tilde{v}_2$ satisfying the given lens aperture D and the ratio between v_1 and v_2, then the estimated depth map $\tilde{s}$ will generate the same blur radii as with the true depth map and the true camera parameters. By Claim 1 when $\tilde{D} = D$ and $\frac{\tilde{v}_1}{\tilde{v}_2} = \frac{v_1}{v_2}$ the estimated shape $\tilde{s}$ satisfies

$$\frac{1}{\tilde{s}(x)} = \frac{v_i}{s(x)\tilde{v}_i} - \frac{1}{\tilde{F}} + \frac{v_i}{F\tilde{v}_i} \quad (20)$$

so that we can immediately verify that for $i = 1, 2$

$$\begin{aligned}\tilde{b}_i(x) &= \frac{\tilde{D}\tilde{v}_i}{2}\left|\frac{1}{\tilde{F}}-\frac{1}{\tilde{v}_i}-\frac{1}{\tilde{s}(x)}\right| \\ &= \frac{D\tilde{v}_i}{2}\left|\frac{1}{\tilde{F}}-\frac{1}{\tilde{v}_i}-\frac{v_i}{s(x)\tilde{v}_i}-\frac{1}{\tilde{F}}+\frac{v_i}{F\tilde{v}_i}\right| \\ &= \frac{Dv_i}{2}\left|\frac{1}{F}-\frac{1}{v_i}-\frac{1}{s(x)}\right| = b_i(x).\end{aligned} \quad (21)$$

Notice that the estimated shape $\tilde{s}$ is still ambiguous, although the blur radii are fully determined. Then, once the blur radii are determined, the PSF h is fully determined and identical to the true PSF. Hence, image restoration can be performed exactly.

Of course, if these two parameters are not known, then image restoration can not be performed exactly unless we introduce additional constraints. We analyze and predict how the radiance will be restored based on partial or no knowledge of these two parameters. We consider the case of unknown D and known ratio $\frac{v_1}{v_2}$. Other cases do not yield any simple analysis of image restoration. If only D is unknown, then the blur radii will be globally and uniformly larger or smaller than the true ones. This means that a simple image restoration algorithm will return a radiance that

is overall oversharpened or undersharpened with respect to the true radiance.

However, in this case it might be possible to perform exact image restoration. If there exist points in the scene that are in focus, it is possible to estimate the aperture diameter D. The knowledge of the ratio $\frac{v_1}{v_2}$ and the reconstructed shape $\tilde{s}$ allow us to predict exactly which points in the scene are in focus for a given setting v_i. Then, since the defocused image with setting v_i coincides with the true radiance at these points, we can use the defocused image with setting v_i as radiance and estimate the exact blur radius in the other defocused image. This will yield the correct aperture diameter D.

5. Experimental validation

In this section we validate the claims proven analytically above on experiments with real and synthetic image sets. For convenience, we have used the data and algorithm for SFD developed by [7] that is publicly available on the web at `http://www.eps.hw.ac.uk/ pf21/downloads`.

5.1. Uncalibrated reconstruction

In this set of experiments we consider 10 synthetic data sets and 2 real data sets. All the input images are shown in Figure 1. In particular, the bottom pairs of images correspond to the real data sets. For each data set we have the true calibration parameters. We test the algorithm for SFD on each data set by selecting calibration parameters that are intentionally different from the ground truth. In Figure 2 we show some examples of reconstructed depth maps when calibration parameters are different from the ground truth. We then match the estimated depth map to the ground truth (the depth map obtained with the correct camera parameters) by using the affine model (12) and compute the corresponding α and β. We compare each estimated parameter with the corresponding parameter predicted by our analysis. The differences between two types of parameters and the results of matching the estimated shapes with the affine model are then listed in Table 1. As one can notice, the low matching errors validate the analysis carried out in the previous sections and show that the affine model derived in eq. (12) holds.

We also show how the matching error changes with noise on the Wave data set. Figure 3 illustrates the matching error (mean and standard deviation) with different levels of white noise, i.e., by adding noise to the original data with variances $0.5, 1, 5, 20, 40$ respectively (in grayscale images with 256 levels). In Figure 3 notice that despite the large amount of noise in the input images (e.g., tick number 5 corresponds to additive Gaussian noise with a variance of 40) , the affine model holds very well.

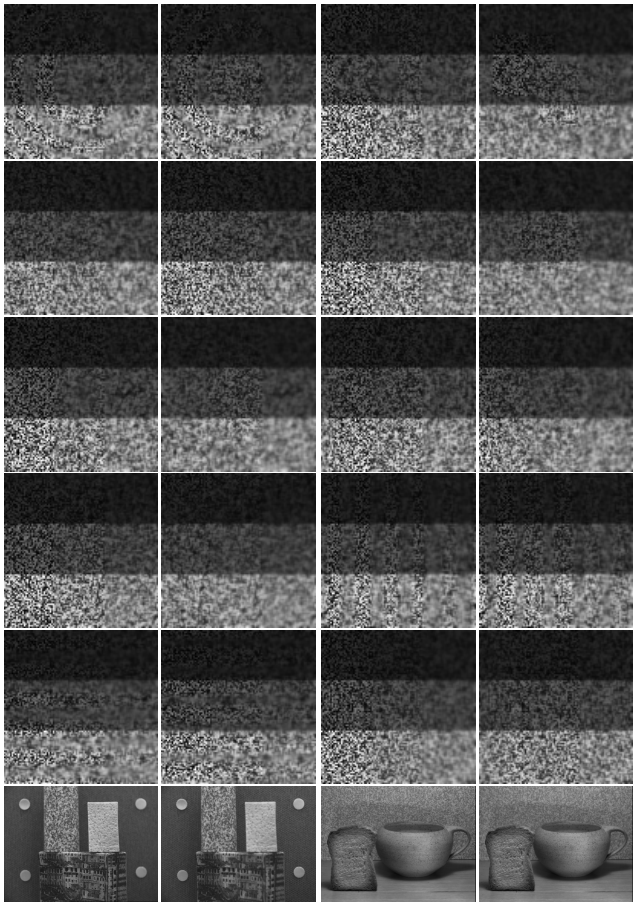

Figure 1. Data sets. The first 5 rows are synthetic data sets while the last row is a real data set. From left to right and top to bottom we have: The Wave data set, the Box data set, the Plane data set, the ExpX data set, the ExpY data set, the InvX data set, the InvY data set, the SinX data set, the SinY data set, the Slope data set, the Cylinder data set and the CupToast data set.

5.2. Autocalibration

In this section, we experimentally verify Claim 2. We show that planarity is not preserved if incorrect calibration parameters are used. To do so, we consider the Slope data set because it is made of a slanted plane not parallel to the image plane. In Figure 4 we display the depth map estimated with correct calibration parameters (blue plane on the left of each image) and compare it with the depth maps estimated with incorrect calibration parameters. In particular, the plane next to the ground truth (the green plane) has been estimated with $\beta = 0$ while the other curved surfaces have been reconstructed with $\beta \neq 0$. On the bottom image of Figure 4 we display the same estimated depth maps from a vantage point that enhances the distortion introduced by incorrect calibration parameters. We compare 3 different experiments with the true depth map.

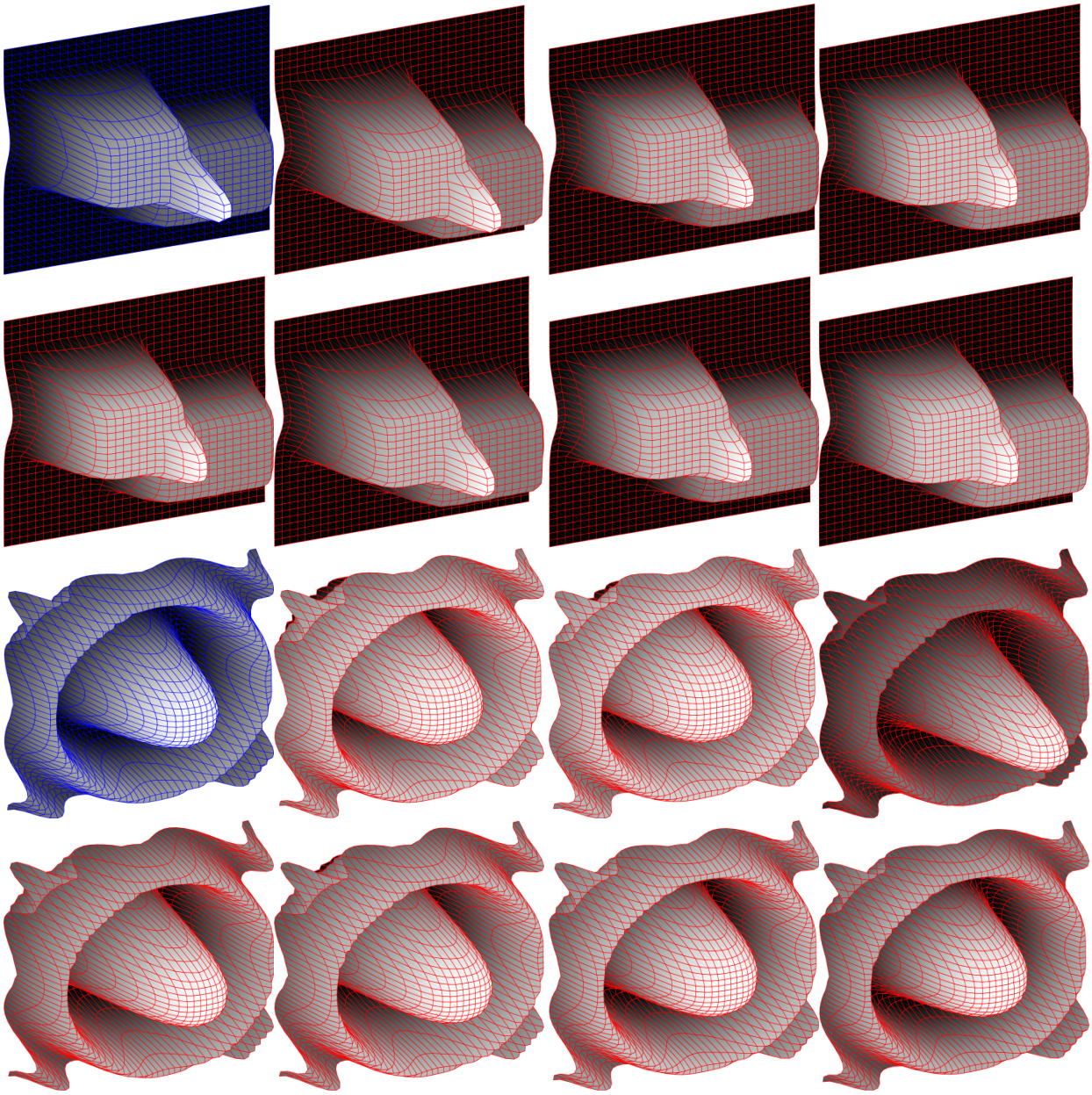

Figure 2. Estimated depth map for the Box and Wave data sets. The top-left image is the depth map of the Box data set estimated with correct calibration parameters; the first image on the left of the third row is the depth map of the Wave data set estimated with correct calibration parameters; all the other images depict depth maps of the corresponding data sets estimated with incorrect camera parameters. Notice that the depth maps estimated with incorrect camera parameters are visually very similar to the true depth map. Indeed, as we show in Table 1, the transformation between each pair of depth maps is the affine model in eq. (12).

5.3. Image restoration

In this section we validate Claim 4. Figure 5 shows that one can reconstruct the radiance without knowledge of the scene depth map. The focal settings and radiance of the synthetic data set "Wave" are known to us. We reconstruct the estimated radiance with random parameters while keeping D and the ratio between v_1 and v_2 unchanged. The relative error between the estimated radiance with random and true parameters is 0.0058 (in grayscale images with 256 levels). In the case of the real data set "CupToast" we compute the estimated radiance both with true and random parameters, yielding 0.0021 as relative error.

Table 1. The means and standard deviations of differences between α and $\hat{\alpha}$, β and $\hat{\beta}$, and matching errors (we display the estimated value multiplied by 10^3), where $\hat{\alpha}, \hat{\beta}$ are obtained by minimizing eq. (12) using least squares, and matching errors are measured by $\|s(x) - \alpha\tilde{s}(x) - \beta s(x)\tilde{s}(x)\|$ in L_2 norm.

	$\lvert\alpha - \hat{\alpha}\rvert$		$\lvert\beta - \hat{\beta}\rvert$		matching errors	
Data set	Mean	Var	Mean	Var	Mean	Var
Box	0.031	0.026	0.059	0.050	0.475	1.167
Plane	0.000	0.000	0.000	0.000	0.000	0.000
ExpX	0.020	0.019	0.038	0.035	0.160	0.287
ExpY	0.045	0.052	0.087	0.099	0.229	0.386
InvX	0.004	0.003	0.006	0.004	0.533	1.189
InvY	0.051	0.031	0.088	0.053	0.972	2.029
SinX	0.001	0.001	0.029	0.027	0.215	0.378
SinY	0.010	0.007	0.036	0.024	0.257	0.275
Slope	0.000	0.000	0.000	0.000	0.000	0.002
Wave	0.186	0.126	0.124	0.084	2.392	3.479
Cylinder	2.323	1.828	4.046	3.184	40.97	108.6
CupToast	0.035	0.021	1.562	0.938	0.193	0.129

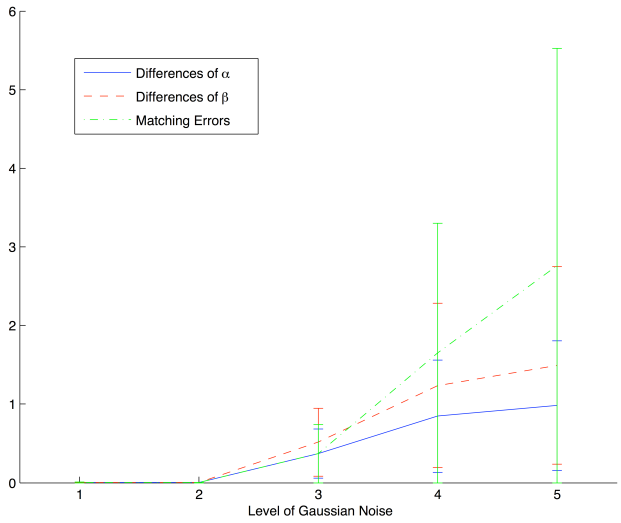

Figure 3. Errors in α, β, and affine model matching as different levels of noise are added to the original images in the Wave Dataset. In the abscissa axis the ticks $1, 2, 3, 4$ and 5 correspond to five levels of Gaussian noise with variance $0.5, 1, 5, 20, 40$ respectively.

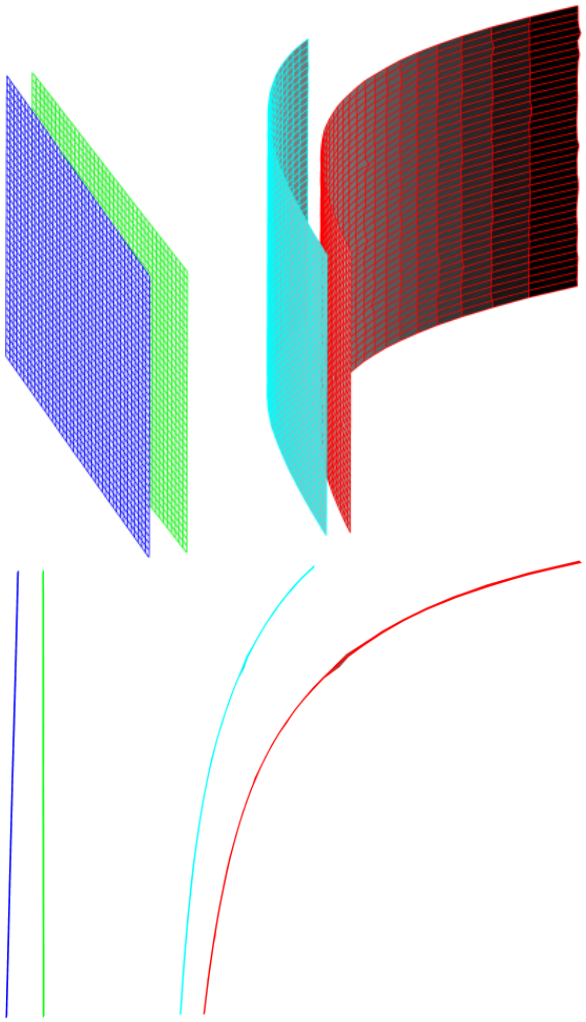

Figure 4. Autocalibration. Planar scenes can be used to calibrate the camera. The top image depicts a side view of the reconstructed depth maps while the bottom image shows the top view of the same plot. On the left of the plot of each image we show the depth map estimated with correct camera parameters (the blue leftmost plane) and the depth map estimated with $\beta = 0$ (the green plane, second from the lef). The two curved surfaces on the right hand side are depth maps that have been estimated with incorrect camera parameters. As one can immediately observe, the depth maps estimated with incorrect camera parameters are no longer planar.

6. Discussion

In this paper we have addressed the problem of recovering shape and radiance of a scene from deblurred images in the absence of calibration information, that is without knowing the lens diameter, focal length and position of the image plane. Following an analogy with structure from motion, we have shown that it is possible to characterize the set of scenes that can be reconstructed from uncalibrated data with a simple analytical model, and that partial knowledge about the scene (e.g., the presence of planes) can be exploited to reduce the ambiguity in uncalibrated reconstruction. Furthermore, we have explored the effects of uncalibrated distortions in the reconstruction on shape in the restoration of the original images, and shown that image restoration is partially sensitive to lack of calibration parameters.

Acknowledgments

This research was supported by AFOSR FA9550-06-1-0138 and ONR N00014-03-1-0850:P0001.

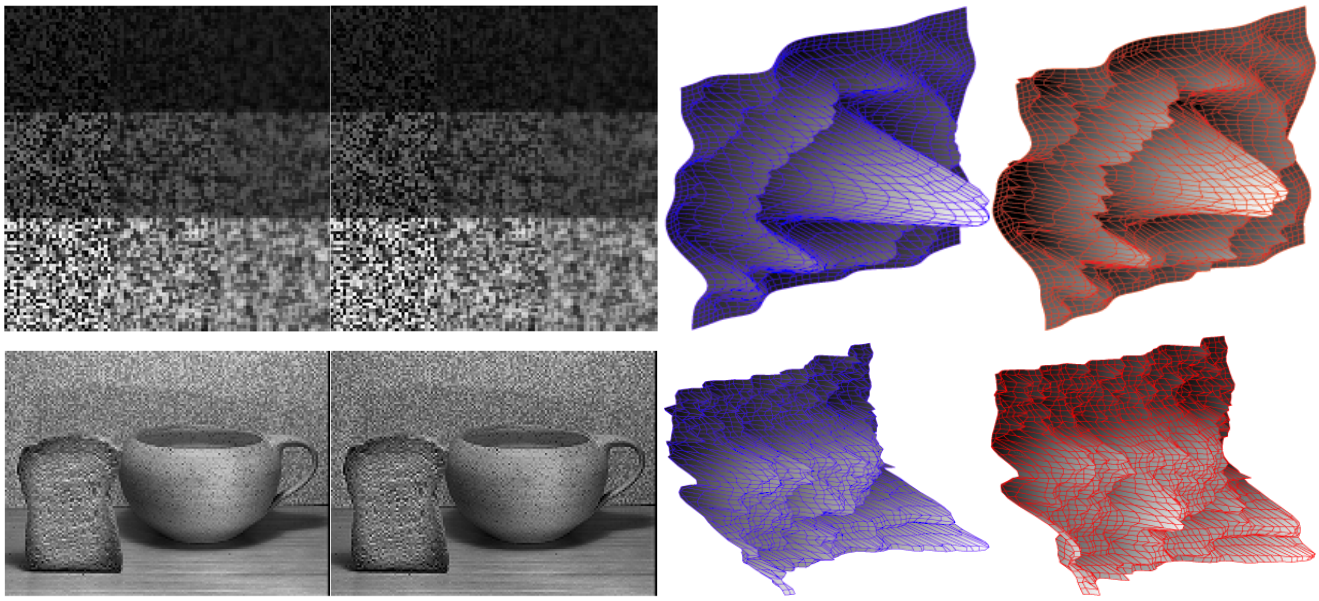

Figure 5. Estimated depth and radiance for the "Wave" and "CupToast" data sets. The top-left image is the true radiance of the Wave data set; the second is the radiance estimated by random parameters with fixed D and $\frac{v_1}{v_2}$. The top-right two images are the true depth map and estimated one respectively. The images in the second row are reconstructed radiance and depth maps for real data set "CupToast" with true and random parameters respectively. Although the depth maps are different, the relative errors of the two radiance functions are small: 0.0058 for the "Wave" data set and 0.0021 for the "CupToast" data set.

References

[1] M. Aggarwal and N. Ahuja. Estimating sensor orientation in cameras. In *ICPR '00: Proceedings of the International Conference on Pattern Recognition*, pages 896–9, Washington, DC, USA, 2000. IEEE Computer Society. 1

[2] M. Baba, M. Mukunoki, and N. Asada. A unified camera calibration using geometry and blur of feature points. In *ICPR '06: Proceedings of the 18th International Conference on Pattern Recognition*, pages 816–9, Washington, DC, USA, 2006. IEEE Computer Society. 1

[3] S. Chaudhuri and A. N. Rajagopalan. Depth from defocus: a real aperture imaging approach. *Springer-Verlag*, 1999. 1, 4

[4] T. Darrell and K. Wohn. Pyramid based depth from focus. In *Computer Vision and Pattern Recognition*, pages 504–509, 1988. 1

[5] O. Faugeras. *Three dimensional vision, a geometric viewpoint.* MIT Press, 1993. 1

[6] P. Favaro, A. Mennucci, and S. Soatto. Observing shape from defocused images. *International Journal of Computer Vision*, 52(1):25–43, 2003. 2

[7] P. Favaro and S. Soatto. A geometric approach to shape from defocus. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(3):406–17, 2005. 1, 5

[8] P. Favaro and S. Soatto. *3-d shape reconstruction and image restoration: exploiting defocus and motion-blur.* Springer-Verlag, 2006. 1

[9] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge University Press, 2000. 1

[10] M. Levoy, B. Chen, V. Vaish, M. Horowitz, I. McDowall, and M. Bolas. Synthetic aperture confocal imaging. *ACM Trans. Graph.*, 23(3):825–834, 2004. 1

[11] Y. Ma, S. Soatto, J. Kosecka, and S. Sastry. *An invitation to 3D vision, from images to models*. Springer Verlag, 2003. 3

[12] A. P. Pentland. A new sense for depth of field. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(4):523–531, July 1987. 1

[13] A. N. Rajagopalan and S. Chaudhuri. Optimal selection of camera parameters for recovery of depth from defocused images. In *Computer Vision and Pattern Recognition*, pages 219–224, 1997. 1

[14] A. N. Rajagopalan and S. Chaudhuri. An mrf model-based approach to simultaneous recovery of depth and restoration from defocused images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(7):577–589, July 1999. 1

[15] A. N. Rajagopalan, S. Chaudhuri, and U. Mudenagudi. Depth estimation and image restoration using defocused stereo pairs. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(11):1521–1525, 2004. 1

[16] R. Tsai. A versatile camera calibration technique for high accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE J. Robotics Automat.*, RA-3(4):323–344, 1987. 1

[17] M. Watanabe and S. K. Nayar. Rational filters for passive depth from defocus. *International Journal of Computer Vision*, 27(3):203–225, May 1998. 1