Multiple-Aperture Photography for High Dynamic Range and Post-Capture Refocusing

Samuel W. Hasinoff, *Member, IEEE,* and Kiriakos N. Kutulakos, *Member, IEEE*

Abstract—In this article we present *multiple-aperture photography*, a new method for analyzing sets of images captured with different aperture settings, with all other camera parameters fixed. Using an image restoration framework, we show that we can simultaneously account for defocus, high dynamic range exposure (HDR), and noise, all of which are confounded according to aperture. Our formulation is based on a layered decomposition of the scene that models occlusion effects in detail. Recovering such a scene representation allows us to adjust the camera parameters in post-capture, to achieve changes in focus setting or depth of field—with all results available in HDR. Our method is designed to work with very few input images: we demonstrate results from real sequences obtained using the three-image "aperture bracketing" mode found on consumer digital SLR cameras.

Index Terms—Computational photography, computer vision, computer graphics, shape-from-defocus, high dynamic range imaging.

✦

1 INTRODUCTION

TYPICAL cameras have three major controls—aperture, shutter speed, and focus. Together, aperture and shutter speed determine the total amount of light incident on the sensor (*i.e.*, exposure), whereas aperture and focus determine the extent of the scene that is in focus (and the degree of out-of-focus blur). Although these controls offer flexibility to the photographer, once an image has been captured, these settings cannot be altered.

Recent computational photography methods aim to free the photographer from this choice by collecting several controlled images [1], [2], [3], or using specialized optics [4], [5]. For example, high dynamic range (HDR) photography involves fusing images taken with varying shutter speed, to recover detail over a wider range of exposures than can be achieved in a single photo [1], [6].

In this article we show that flexibility can be greatly increased through multiple-aperture photography, *i.e.*, by collecting several images of the scene with all settings except aperture fixed (Fig. 1). In particular, our method is designed to work with very few input images, including the three-image "aperture bracketing" mode found on most consumer digital SLR cameras. Multiple-aperture photography takes advantage of the fact that by controlling aperture we simultaneously modify the exposure and defocus of the scene. To our knowledge, defocus has not previously been considered in the context of widely-ranging exposures.

We show that by inverting the image formation in

the input photos, we can decouple all three controls—aperture, focus, and exposure—thereby allowing complete freedom in post-capture, *i.e.*, we can resynthesize HDR images for any user-specified focus position or aperture setting. While this is the major strength of our technique, it also presents a significant technical challenge. To address this challenge, we pose the problem in an image restoration framework, connecting the radiometric effects of the lens, the depth and radiance of the scene, and the defocus induced by aperture.

The key to the success of our approach is formulating an image formation model that accurately accounts for the input images, and allows the resulting image restoration problem to be inverted in a tractable way, with gradients that can be computed analytically. By applying the image formation model in the forward direction we can resynthesize images with arbitrary camera settings, and even extrapolate beyond the settings of the input.

In our formulation, the scene is represented in layered form, but we take care to model occlusion effects at defocused layer boundaries [7] in a physically meaningful way. Though several depth-from-defocus methods have previously addressed such occlusion, these methods have been limited by computational inefficiency [8], a restrictive occlusion model [9], or the assumption that the scene is composed of two surfaces [8], [9], [10]. By comparison, our approach can handle an arbitrary number of layers, and incorporates an approximation that is effective and efficient to compute. Like McGuire, *et al.* [10], we formulate our image formation model in terms of image compositing [11], however our analysis is not limited to a two-layer scene or input photos with special focus settings.

Our work is also closely related to depth-from-defocus methods based on image restoration, that recover an all-in-focus representation of the scene [8], [12], [13], [14]. Although the output of these methods theoretically

- *S. W. Hasinoff is with the Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02140. Email: hasinoff@csail.mit.edu.*

- *K. N. Kutulakos is with the Department of Computer Science, University of Toronto, Canada M5S 3G4. Email: kyros@cs.toronto.edu.*

multiple-aperture input photos

f/8   f/4   f/2

post-capture resynthesis, in HDR

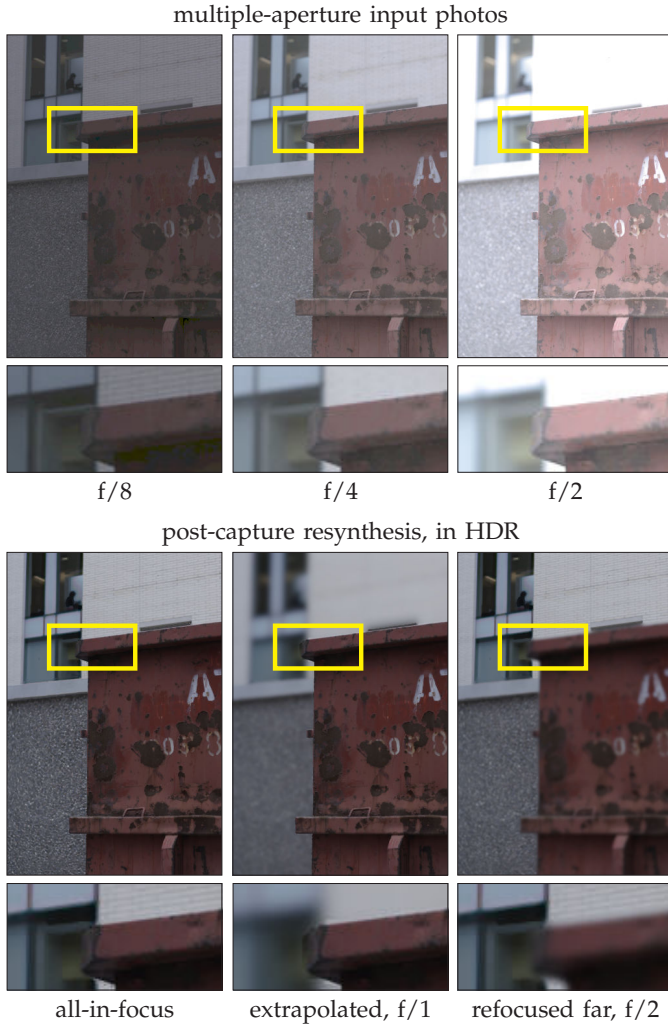all-in-focus   extrapolated, f/1   refocused far, f/2

Fig. 1. Photography with varying apertures. *Top:* Input photographs for the DUMPSTER dataset, obtained by varying aperture setting only. Without the strong gamma correction we apply for display ($\gamma = 3$), these images would appear extremely dark or bright, since they span a wide exposure range. Note that aperture affects both exposure and defocus. *Bottom:* Examples of post-capture resynthesis, shown in high dynamic range (HDR) with tone-mapping. Left-to-right: the all-in-focus image, an extrapolated aperture (f/1), and refocusing on the background (f/2).

permits post-capture refocusing and aperture control, most of these methods assume an additive, transparent image formation model [12], [13], [14] which causes serious artifacts at depth discontinuities, due to the lack of occlusion modeling. Similarly, defocus-based techniques specifically designed to allow refocusing rely on inverse filtering with local windows [15], [16], and do not model occlusion either. Importantly, none of these methods are designed to handle the large exposure differences found in multiple-aperture photography.

Our work has four main contributions. First, we introduce multiple-aperture photography as a way to decou-

ple exposure and defocus from a sequence of images. Second, we propose a layered image formation model that is efficient to evaluate, and enables accurate resynthesis by accounting for occlusion at defocused boundaries. Third, we show that this formulation is specifically designed for an objective function that can be practicably optimized within a standard restoration framework. Fourth, as our experimental results demonstrate, multiple-aperture photography allows post-capture manipulation of all three camera controls—aperture, shutter speed, and focus—from the same number of images used in basic HDR photography.

2 PHOTOGRAPHY BY VARYING APERTURE

Suppose we have a set of photographs of a scene taken from the same viewpoint with different apertures, holding all other camera settings fixed. Under this scenario, image formation can be expressed in terms of four components: a scene-independent lens attenuation factor $\mathbf{R}$, a scene radiance term $\overline{\mathbf{L}}$, the sensor response function $g(\cdot)$, and image noise η,

$$
\mathbf{I}(x,y,a) = g\Big(\overbrace{\underbrace{\mathbf{R}(x,y,a,f)}_{\text{lens term}} \cdot \underbrace{\overline{\mathbf{L}}(x,y,a,f)}_{\text{scene radiance term}}}^{\text{sensor irradiance}} \Big) + \underbrace{\eta}_{\text{noise}} ,
$$
(1)

where $\mathbf{I}(x,y,a)$ is image intensity at pixel (x,y) when the aperture is a. In this expression, the lens term $\mathbf{R}$ models the radiometric effects of the lens and depends on pixel position, aperture, and the focus setting, f, of the lens. The radiance term $\overline{\mathbf{L}}$ corresponds to the mean scene radiance integrated over the aperture, *i.e.*, the total radiance subtended by aperture a divided by the solid angle. We use mean radiance because this allows us to decouple the effects of exposure, which depends on aperture but is scene-independent, and of defocus, which also depends on aperture.

Given the set of captured images, our goal is to perform two operations:

- **High dynamic range photography.** Convert each of the input photos to HDR, *i.e.*, recover $\overline{\mathbf{L}}(x,y,a,f)$ for the input camera settings, (a,f).
- **Post-capture aperture and focus control.** Compute $\overline{\mathbf{L}}(x,y,a',f')$ for any aperture and focus setting, (a',f').

Computing an HDR photograph from images where exposure time is the only control is relatively straightforward because exposure time only affects the brightness of each pixel. In contrast, in our approach, where aperture varies across photos, defocus and exposure are deeply interrelated. Hence, existing HDR and defocus analysis methods do not apply, and an entirely new inverse problem must be formulated and solved.

To do this, we establish a computationally tractable model for the terms in Eq. (1) that approximates well the image formation in off-the-shelf digital cameras. Importantly, we show that this model leads to a restoration-
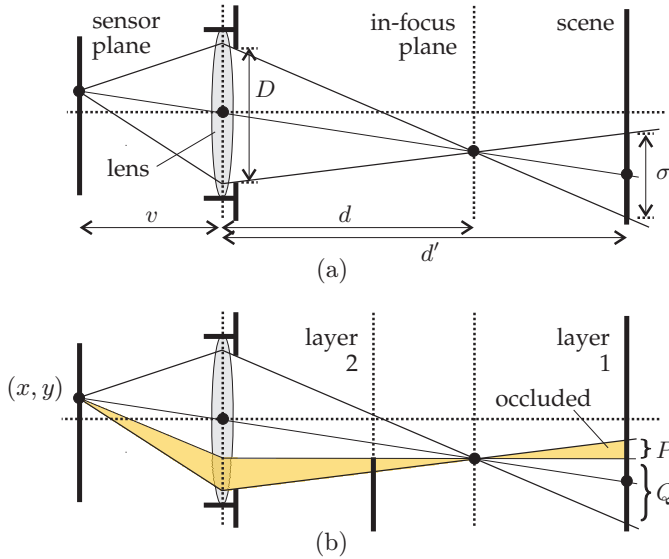
Fig. 2. Defocused image formation with the thin lens model. (a) Fronto-parallel scene. (b) For a two-layered scene, the shaded fraction of the cone integrates radiance from layer 2 only, while the unshaded fraction integrates the unoccluded part of layer 1. Our occlusion model of Sec. 4 approximates layer 1's contribution to the radiance at (x, y) as $(\mathbf{L}_P + \mathbf{L}_Q)\frac{|Q|}{|P|+|Q|}$, where $\mathbf{L}_P$ and $\mathbf{L}_Q$ represent the total radiance from regions P and Q respectively. This is a good approximation when $\frac{1}{|P|}\mathbf{L}_P \approx \frac{1}{|Q|}\mathbf{L}_Q$.

based optimization problem that can be solved efficiently.

3 IMAGE FORMATION MODEL

Sensor model. Following the HDR photography literature [1], we express the sensor response $g(\cdot)$ in Eq. (1) as a smooth, monotonic function mapping the sensor irradiance $\mathbf{R} \cdot \overline{\mathbf{L}}$ to image intensity in the range $[0, 1]$. The effective dynamic range is limited by over-saturation, quantization, and the sensor noise η, which we model as additive.

Exposure model. Since we hold exposure time constant, a key factor in determining the magnitude of sensor irradiance is the size of the aperture. In particular, we represent the total solid angle subtended by the aperture with an exposure factor e_a that maps the mean radiance, $\overline{\mathbf{L}}$, to the total radiance integrated over the aperture, $e_a\overline{\mathbf{L}}$. Because this factor is scene-independent, we incorporate it in the lens term,

$$\mathbf{R}(x, y, a, f) = e_a \, \hat{\mathbf{R}}(x, y, a, f) \quad , \tag{2}$$

therefore the factor $\hat{\mathbf{R}}(x, y, a, f)$ models residual radiometric distortions, such as vignetting [17], that vary spatially and depend on aperture and focus setting. To resolve the multiplicative ambiguity, we assume that $\hat{\mathbf{R}}$ is normalized so the center pixel is assigned a factor of one.

Defocus model. While more general models are possible [18], we assume that the defocus induced by the aperture obeys the standard thin lens model [7], [19]. This model has the attractive feature that for a fronto-parallel scene, relative changes in defocus due to aperture setting are independent of depth.

In particular, for a fronto-parallel scene with radiance $\mathbf{L}$, the defocus from a given aperture can be expressed by the convolution $\overline{\mathbf{L}} = \mathbf{L} * B_\sigma$ [19]. The 2D point-spread function B is parameterized by the effective *blur diameter*, σ, which depends on scene depth, focus setting, and aperture size (Fig. 2a). From simple geometry,

$$\sigma = \frac{|d' - d|}{d}D \quad , \tag{3}$$

where d' is the depth of the scene, d is the depth of the in-focus plane, and D is the effective diameter of the aperture. This implies that regardless of the scene depth, for a fixed focus setting, the blur diameter is proportional to the aperture diameter.[1]

The thin lens geometry also implies that whatever its form, the point-spread function B will scale radially with blur diameter, *i.e.*, $B_\sigma(x, y) = \frac{1}{\sigma^2}B(\frac{x}{\sigma}, \frac{y}{\sigma})$. In practice, we assume that B_σ is a 2D symmetric Gaussian, where σ represents the standard deviation of the point-spread function, $B_\sigma(x, y) = \frac{1}{2\pi\sigma^2}e^{-(x^2+y^2)/2\sigma^2}$.

4 LAYERED SCENE RADIANCE

To make the reconstruction problem tractable, we rely on a simplified scene model that consists of multiple, possibly overlapping, fronto-parallel layers, ideally corresponding to a gross object-level segmentation of the 3D scene.

In this model, the scene is composed of K layers, numbered from back to front. Each layer is specified by an HDR image, $\mathbf{L}_k$, that describes its outgoing radiance at each point, and an alpha matte, $\mathbf{A}_k$, that describes its spatial extent and transparency.

4.1 Approximate layered occlusion model

Although the relationship between defocus and aperture setting is particularly simple for a single-layer scene, the multiple layer case is significantly more challenging due to occlusion.[2] A fully accurate simulation of the thin lens model under occlusion involves backprojecting a cone into the scene, and integrating the unoccluded radiance (Fig. 2b) using a form of ray-tracing [7]. Unfortunately, this process is computationally intensive, since the point-spread function can vary with arbitrary complexity according to the geometry of the occlusion boundaries.

---

1. Because it is based on simple convolution, the thin lens model for defocus implicitly assumes that scene radiance $\mathbf{L}$ is constant over the cone subtended by the largest aperture [20], [21]. The model also implies that any camera settings yielding the same blur diameter σ will produce the same defocused image.

2. Since we model the layers as thin in depth, occlusion due to surfaces that are parallel to the optical axis [9] can be ignored.
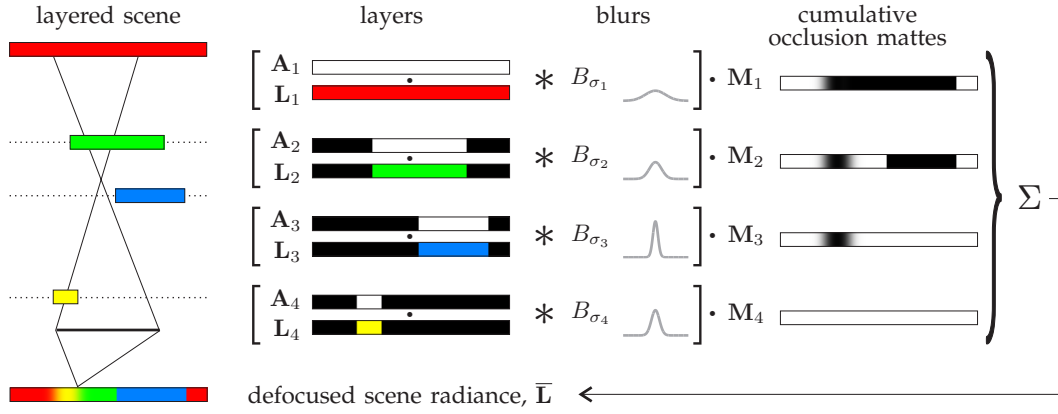
Fig. 3. Approximate layered image formation model with occlusion, illustrated in 2D. The double-cone shows the thin lens geometry for a given pixel, indicating that layer 3 is nearly in-focus. To compute the defocused radiance, $\overline{\mathbf{L}}$, we use convolution to independently defocus each layer $\mathbf{A}_k \cdot \mathbf{L}_k$, where the blur diameters σ_k are defined by the depths of the layers (Eq. (3)). We combine the independently defocused layers using image compositing, where the mattes $\mathbf{M}_k$ account for cumulative occlusion from defocused layers in front.

For computational efficiency, we therefore formulate an approximate model for layered image formation (Fig. 3) that accounts for occlusion, is effective in practice, and leads to simple analytic gradients used for optimization.

The model entails defocusing each scene layer independently, according to its depth, and combining the results using image compositing:

$$\overline{\mathbf{L}} = \sum_{k=1}^{K} \left[(\mathbf{A}_k \cdot \mathbf{L}_k) * B_{\sigma_k} \right] \cdot \mathbf{M}_k \quad , \tag{4}$$

where σ_k is the blur diameter for layer k, $\mathbf{M}_k$ is a second alpha matte for layer k, representing the cumulative occlusion from defocused layers in front,

$$\mathbf{M}_k = \prod_{j=k+1}^{K} \left(1 - \mathbf{A}_j * B_{\sigma_j} \right) \quad , \tag{5}$$

and $\cdot$ denotes pixel-wise multiplication. Eqs. (4) and (5) can be viewed as an application of the matting equation [11], and generalizes the method of McGuire, *et al.* [10] to arbitrary focus settings and numbers of layers.

Intuitively, rather than integrating partial cones of rays that are restricted by the geometry of the occlusion boundaries (Fig. 2b), we integrate the entire cone for each layer, and weigh each layer's contribution by the fraction of rays that reach it. These weights are given by the alpha mattes, and model the thin lens geometry exactly.

In general, our approximation is accurate when the region of a layer that is subtended by the entire aperture has the same mean radiance as the unoccluded region (Fig. 2b). This assumption is less accurate when only a small fraction of the layer is unoccluded, but this case is mitigated by the small contribution of the layer to the overall integral. Worst-case behavior occurs when an occlusion boundary is accidentally aligned with a brightness or texture discontinuity on the occluded layer, however this is rare in practice.

4.2 All-in-focus scene representation

In order to simplify our formulation even further, we represent the entire scene as a single all-in-focus HDR radiance map, $\mathbf{L}$. In this reduced representation, each layer is modeled as a binary alpha matte $\mathbf{A}'_k$ that "selects" the unoccluded pixels corresponding to that layer. Note that if the narrowest-aperture input photo is all-in-focus, the brightest regions of $\mathbf{L}$ can be recovered directly, however this condition is not a requirement of our method.

While the all-in-focus radiance directly specifies the unoccluded radiance $\mathbf{A}'_k \cdot \mathbf{L}$ for each layer, to accurately model defocus near layer boundaries we must also estimate the radiance for occluded regions (Fig. 2b). Our underlying assumption is that $\mathbf{L}$ is sufficient to describe these occluded regions as extensions of the unoccluded layers. This allows us to apply the same image formation model of Eqs. (4)–(5) to extended versions of the unoccluded layers (Fig. 4):

$$\mathbf{A}_k = \mathbf{A}'_k \quad + \mathbf{A}''_k \tag{6}$$
$$\mathbf{L}_k = \mathbf{A}'_k \cdot \mathbf{L} + \mathbf{A}''_k \cdot \mathbf{L}''_k \quad . \tag{7}$$

In Sec. 7 we describe our method for extending the unoccluded layers using image inpainting.

4.3 Complete scene model

In summary, we represent the scene by the triple $(\mathbf{L}, \mathbf{A}, \sigma)$, consisting of the all-in-focus HDR scene radiance, $\mathbf{L}$, the hard segmentation of the scene into unoccluded layers, $\mathbf{A} = \{\mathbf{A}'_k\}$, and the per-layer blur diameters, σ, specified for the widest aperture.[3]

---

3. To relate the blur diameters over aperture setting, we rely on Eq. (3). Note that in practice we do not compute the aperture diameters directly from the f-numbers. For greater accuracy, we instead estimate the relative aperture diameters according to the calibrated exposure factors, $D_a \propto \sqrt{e_a/e_A}$.
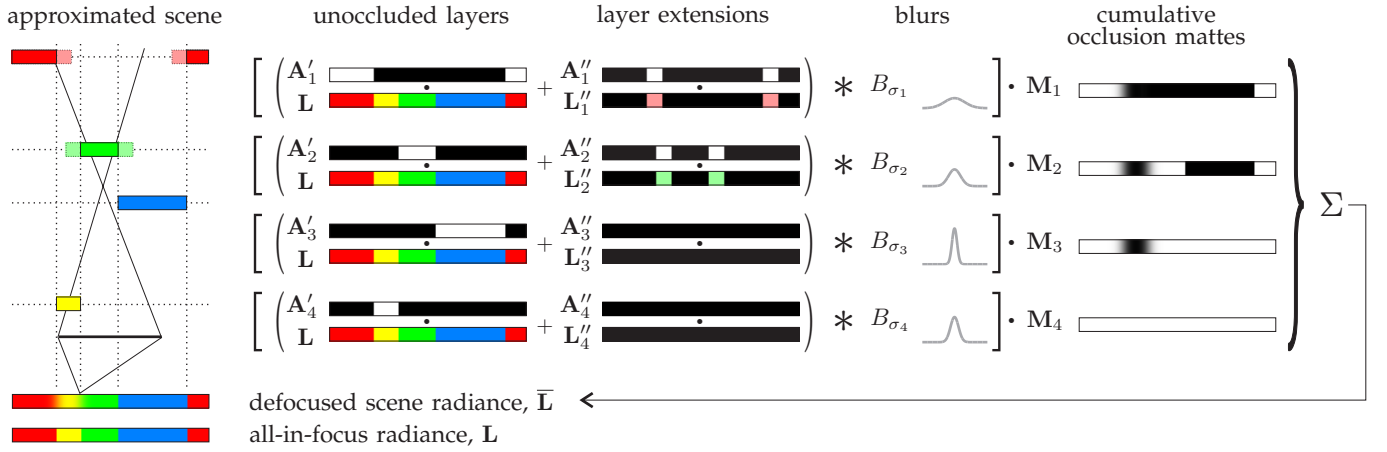
Fig. 4. Reduced representation for the layered scene in Fig. 3, based on the all-in-focus radiance, $\mathbf{L}$. The all-in-focus radiance specifies the unoccluded regions of each layer, $\mathbf{A}'_k \cdot \mathbf{L}$, where $\{\mathbf{A}'_k\}$ is a hard segmentation of the unoccluded radiance into layers. We assume that $\mathbf{L}$ is sufficient to describe the occluded regions of the scene as well, with inpainting (lighter, dotted) used to extend the unoccluded regions behind occluders as required. Given these extended layers, $\mathbf{A}'_k \cdot \mathbf{L} + \mathbf{A}''_k \cdot \mathbf{L}''_k$, we apply the same image formation model as in Fig. 3.

5 RESTORATION-BASED FRAMEWORK FOR HDR LAYER DECOMPOSITION

In multiple-aperture photography we do not have any prior information about either the layer decomposition (*i.e.*, depth) or scene radiance. We therefore formulate an inverse problem whose goal is to compute $(\mathbf{L}, \mathbf{A}, \sigma)$ from a set of input photos. The resulting optimization can be viewed as a generalized image restoration problem that unifies HDR imaging and depth-from-defocus by jointly explaining the input in terms of layered HDR radiance, exposure, and defocus.

In particular we formulate our goal as estimating $(\mathbf{L}, \mathbf{A}, \sigma)$ that best reproduces the input images, by minimizing the objective function

$$\mathcal{O}(\mathbf{L}, \mathbf{A}, \sigma) = \frac{1}{2}\sum_{a=1}^{A}\|\mathbf{\Delta}(x,y,a)\|^2 + \lambda\|\mathbf{L}\|_\beta \ . \quad (8)$$

In this optimization, $\mathbf{\Delta}(x,y,a)$ is the residual pixel-wise error between each input image $\mathbf{I}(x,y,a)$ and the corresponding synthesized image; $\|\mathbf{L}\|_\beta$ is a regularization term that favors piecewise smooth scene radiance; and $\lambda > 0$ controls the balance between squared image error and the regularization term.

The following equation shows the complete expression for the residual $\mathbf{\Delta}(x,y,a)$, parsed into simpler components: The residual is defined in terms of input images that have been linearized and lens-corrected according to pre-calibration (Sec. 7). This transformation simplifies the optimization of Eq. (8), and converts the image formation model of Eq. (1) to scaling by an exposure factor e_a, followed by clipping to model over-saturation. The innermost component of Eq. (10) is the layered image formation model described in Sec. 4.

While scaling due to the exposure factor greatly affects the relative magnitude of the additive noise, η, this effect is handled implicitly by the restoration. Note, however, that additive noise from Eq. (1) is modulated by the linearizing transformation that we apply to the input images, yielding modified additive noise at every pixel:

$$\eta'(x,y,a) = \frac{1}{\hat{\mathbf{R}}(x,y,a,f)}\left|\frac{\mathrm{d}g^{-1}(\mathbf{I}(x,y))}{\mathrm{d}\mathbf{I}(x,y)}\right|\eta \ , \quad (11)$$

where $\eta' \to \infty$ for over-saturated pixels [22].

5.1 Weighted TV regularization

To regularize Eq. (8), we use a form of the total variation (TV) norm, $\|\mathbf{L}\|_{\mathrm{TV}} = \int\|\nabla\mathbf{L}\|$. This norm is useful for restoring sharp discontinuities, while suppressing noise and other high frequency detail [23]. The variant we propose,

$$\|\mathbf{L}\|_\beta = \int\sqrt{\left(w(\mathbf{L})\,\|\nabla\mathbf{L}\|\right)^2 + \beta} \ , \quad (12)$$

includes a perturbation term $\beta > 0$ that remains constant[4] and ensures differentiability as $\nabla\mathbf{L} \to 0$ [23]. More importantly, our norm incorporates per-pixel weights $w(\mathbf{L})$ meant to equalize the TV penalty over the high dynamic range of scene radiance (Fig. 12).

We define the weight $w(\mathbf{L})$ for each pixel according to its inverse exposure level, $1/e_{a^*}$, where a^* corresponds to the aperture for which the pixel is "best exposed". In particular, we synthesize the transformed input images using the current scene estimate, and for each pixel we select the aperture with highest signal-to-noise ratio, computed with the noise level η' predicted by Eq. (11).

6 OPTIMIZATION METHOD

To optimize Eq. (8), we use a series of alternating minimizations, each of which estimates one of $\mathbf{L}, \mathbf{A}, \sigma$ while holding the rest constant.

---

4. We used $\beta = 10^{-8}$ in all our experiments.

$$\boldsymbol{\Delta}(x,y,a) = \underbrace{\frac{1}{\hat{\mathbf{R}}(x,y,a,f)}\, g^{-1}\big(\mathbf{I}(x,y,a)\big)}_{\substack{\text{linearized and lens-corrected}\\\text{image intensity}}} - \min\left\{ \underbrace{e_a}_{\substack{\text{exposure}\\\text{factor}}} \cdot \underbrace{\left[\sum_{k=1}^{K}\big[(\mathbf{A}_k\mathbf{L} + \mathbf{A}_k^*\mathbf{L}_k^*) * B_{\sigma_{a,k}}\big]\cdot\mathbf{M}_k\right]}_{\substack{\text{layered occlusion model}\\\text{from Eqs. (4)-(5)}}}, \underbrace{1}_{\substack{\text{clipping}\\\text{term}}} \right\}, \qquad (10)$$

- **Image restoration** To recover the scene radiance **L** that minimizes the objective, we take a direct iterative approach [14], [23], by carrying out a set of conjugate gradient steps. Our formulation ensures that the required gradients have straightforward analytic formulas (Appendix A).
- **Blur refinement** We use the same approach, of taking conjugate gradient steps, to optimize the blur diameters σ. Again, the required gradients have simple analytic formulas (Appendix A).
- **Layer refinement** The layer decomposition **A** is more challenging to optimize because it involves a discrete labeling, but efficient optimization methods such as graph cuts [24] are not applicable. We use a naïve approach that simultaneously modifies the layer assignment of all pixels whose residual error is more than five times the median, until convergence. Each iteration in this stage evaluates whether a change in the pixels' layer assignment leads to a reduction in the objective.
- **Layer ordering** Recall that the indexing for **A** specifies the depth ordering of the layers, from back to front. To test modifications to this ordering, we note that each blur diameter corresponds to two possible depths, either in front of or behind the in-focus plane (Eq. (3)). We use a brute force approach that tests all 2^{K-1} distinct layer orderings, and select the one leading to the lowest objective (Fig. 6d).

  Note that even when the layer ordering and blur diameters are specified, a two-fold ambiguity still remains. In particular, our defocus model alone does not let us resolve whether the layer with the smallest blur diameter (*i.e.*, the most in-focus layer) is in front of or behind the in-focus plane. In terms of resynthesizing new images, this ambiguity has little impact provided that the layer with the smallest blur diameter is nearly in focus. For greater levels of defocus, however, the ambiguity can be significant. Our current approach is to break the ambiguity arbitrarily, but we could potentially analyze errors at occlusion boundaries or exploit additional information (*e.g.*, that the lens is focused behind the scene [25]) to resolve this.
- **Initialization** In order for this procedure to work, we need to initialize all three of $(\mathbf{L}, \mathbf{A}, \sigma)$ with reasonable estimates, as discussed below.

7 IMPLEMENTATION DETAILS

Scene radiance initialization. We define an initial estimate for the unoccluded radiance, **L**, by directly selecting pixels from the transformed input images, then
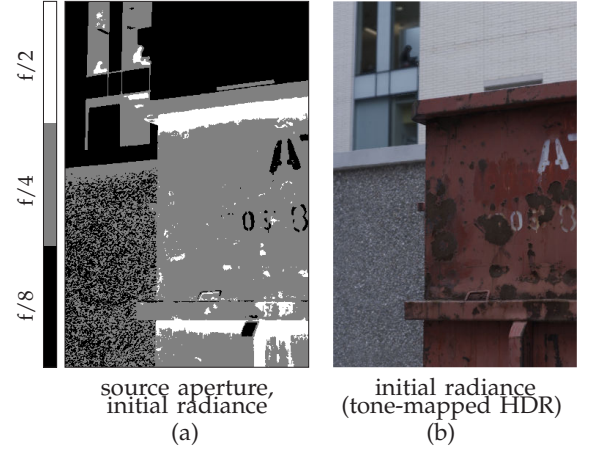

Fig. 5. Initial estimate for unoccluded scene radiance. (a) Source aperture from the input sequence, corresponding to the narrowest aperture with acceptable SNR. (b) Initial estimate for HDR scene radiance, shown using tone-mapping.

scaling them by their inverse exposure factor, $1/e_a$, to convert them to HDR radiance. Our strategy is to select as many pixels as possible from the sharply focused narrowest-aperture image, but to make adjustments for darker regions of the scene, whose narrow-aperture image intensities will be dominated by noise (Fig. 5).

For each pixel, we select the narrowest aperture for which the image intensity is above a fixed threshold of $\kappa = 0.1$, or if none meet this threshold, then we select the largest aperture. In terms of Eq. (11), the threshold defines a minimum acceptable signal-to-noise ratio of κ/η'.

Initial layering and blur assignment. To obtain an initial estimate for the layers and blur diameters, we use a simple window-based depth-from-defocus method inspired by classic approaches [16], [19] and more recent MRF-based techniques [3], [12]. Our method involves directly testing a set of hypotheses for blur diameter, $\{\hat{\sigma}_i\}$, by synthetically defocusing the image as if the whole scene were a single fronto-parallel surface. We specify these hypotheses for blur diameter in the widest aperture, recalling that Eq. (3) relates each such hypothesis over all aperture settings.

Because of the large exposure differences between photos taken several f-stops apart, we restrict our evaluation of consistency with a given blur hypothesis, $\hat{\sigma}_i$, to adjacent pairs of images captured with successive aperture settings, $(a, a+1)$.

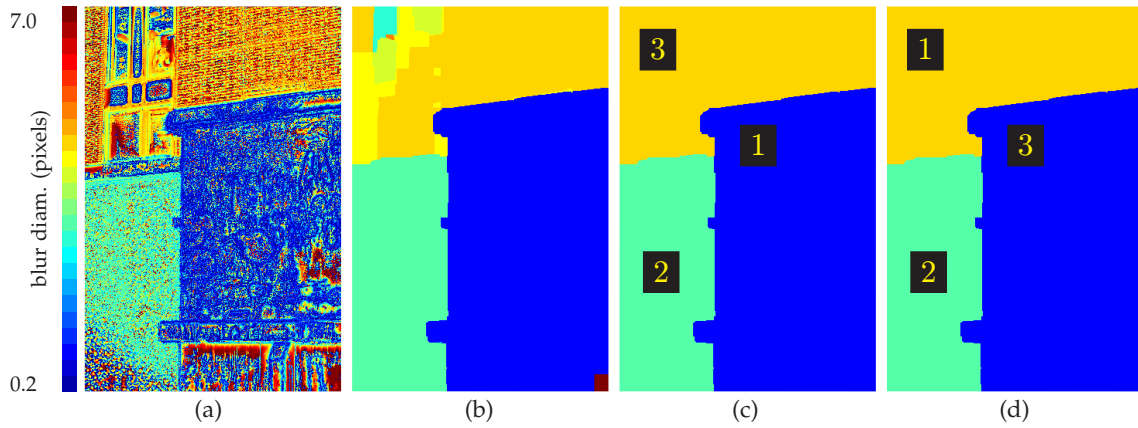To evaluate consistency for each such pair, we use

Fig. 6. (a)–(c) Initial layer decomposition and blur assignment for the DUMPSTER dataset, computed using our depth-from-defocus method. (a) Greedy layer assignment. (b) MRF-based layer assignment. (c) Initial layer decomposition, determined by applying morphological post-processing to (b). Our initial guess for the back-to-front depth ordering is also shown. (d) Final layering, which involves re-estimating the depth ordering and iteratively modifying the layer assignment for high-residual pixels. The corrected depth ordering significantly improves the quality of resynthesis, however the effect of modifying the layer assignment is very subtle.

the hypothesis to align the narrower aperture image to the wider one, then directly measure per-pixel resynthesis error. This alignment involves convolving the narrower aperture image with the required incremental blur, scaling the image intensity by a factor of e_{a+1}/e_a, and clipping any oversaturated pixels. Since our point-spread function is Gaussian, this incremental blur can be expressed in a particularly simple form, namely another 2D symmetric Gaussian with a standard deviation of $(D_{a+1}^2 - D_a^2)^{\frac{1}{2}} \hat{\sigma}_i$.

By summing the resynthesis error across all adjacent pairs of apertures, we obtain a rough per-pixel metric describing consistency with the input images over our set of blur diameter hypotheses. While this error metric can be minimized in a greedy fashion for every pixel (Fig. 6a), we a use Markov random field (MRF) framework to reward piecewise smoothness and recover a small number of layers (Fig. 6b). In particular, we employ graph cuts with the expansion-move approach [26], where the smoothness cost is defined as a truncated linear function of adjacent label differences on the four-connected grid,

$$\sum_{(x',y') \in \text{neigh}(x,y)} \max \left\{ |l(x',y') - l(x,y)|, s_{\max} \right\} , \quad (13)$$

where $l(x,y)$ represents the discrete index of the blur hypothesis $\hat{\sigma}_i$ assigned to pixel (x,y), and $\text{neigh}(x,y)$ defines the adjacency structure. In all our experiments we used $s_{\max} = 2$.

After finding the MRF solution, we apply simple morphological post-processing to detect pixels belonging to very small regions, constituting less than 5% of the image area, and to relabel them according to their nearest neighboring region above this size threshold. Note that our implementation currently assumes that all pixels assigned to the same blur hypothesis belong the

same depth layer. While this simplifying assumption is appropriate for all our examples (*e.g.*, the two window panes in Fig. 14) and limits the number of layers, a more general approach is to assign disconnected regions of pixels to separate layers (we did not do this in our implementation).

Sensor response and lens term calibration. To recover the sensor response function, $g(\cdot)$, we apply standard HDR imaging methods [1] to a calibration sequence captured with varying exposure time.

We recover the radiometric lens term $\mathbf{R}(x, y, a, f)$ using one-time pre-calibration process as well. To do this, we capture a calibration sequence of a diffuse and textureless plane, and compute the radiometric term on a per-pixel basis using simple ratios [20]. In practice our implementation ignores the dependence of $\mathbf{R}$ on focus setting, but if the focus setting is recorded at capture time, we can use it to interpolate over a more detailed radiometric calibration measured over a range of focus settings [20].

Occluded radiance estimation. As illustrated in Fig. 4, we assume that all scene layers can be expressed in terms of the unoccluded all-in-focus radiance $\mathbf{L}$. During optimization, we use a simple inpainting method to extend the unoccluded layers: we use a naïve, low-cost technique that extends each layer by filling its occluded background with the closest unoccluded pixel from its boundary (Fig. 7b). For synthesis, however, we obtain higher-quality results by using a simple variant of PDE-based inpainting [27] (Fig. 7c), which formulates inpainting as a diffusion process. Previous approaches have used similar inpainting methods for synthesis [10], [28], and have also explored using texture synthesis to extend the unoccluded layers [29].

masked bg   inpainted bg   inpainted bg
                (nearest pixel)   (diffusion)
(a)   (b)   (c)

Fig. 7. Layering and background inpainting for the DUMP-STER dataset. (a) The three recovered scene layers, visualized by masking out the background. (b) Inpainting the background for each layer using the nearest layer pixel. (c) Using diffusion-based inpainting [27] to define the layer background. In practice, we need not compute the inpainting for the front-most layer (bottom row).

8   RESULTS AND DISCUSSION

To evaluate our approach we captured several real datasets using two different digital SLR cameras. We also generated a synthetic dataset to enable comparison with ground truth (LENA dataset).

We captured the real datasets using the Canon EOS-1Ds Mark II (DUMPSTER, PORTRAIT, MACRO datasets) or the EOS-1Ds Mark III (DOORS dataset), secured on a sturdy tripod. In both cases we used a wide-aperture fixed focal length lens, the Canon EF85mm f1.2L and the EF50mm f1.2L respectively, set to manual focus. For all our experiments we used the built-in three-image "aperture bracketing" mode set to ±2 stops, and chose the shutter speed so that the images were captured at f/8, f/4, and f/2 (yielding relative exposure levels of roughly 1, 4, and 16). We captured 14-bit RAW images for increased dynamic range, and demonstrate our method for downsampled images with resolutions of 500×333
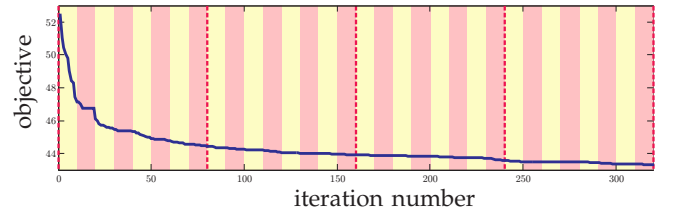

Fig. 8. Typical convergence behavior of our restoration method, shown for the DUMPSTER dataset (Fig. 1). The yellow and pink shaded regions correspond to alternating blocks of image restoration and blur refinement respectively (10 iterations each), and the dashed red vertical lines indicate layer reordering and refinement (every 80 iterations).
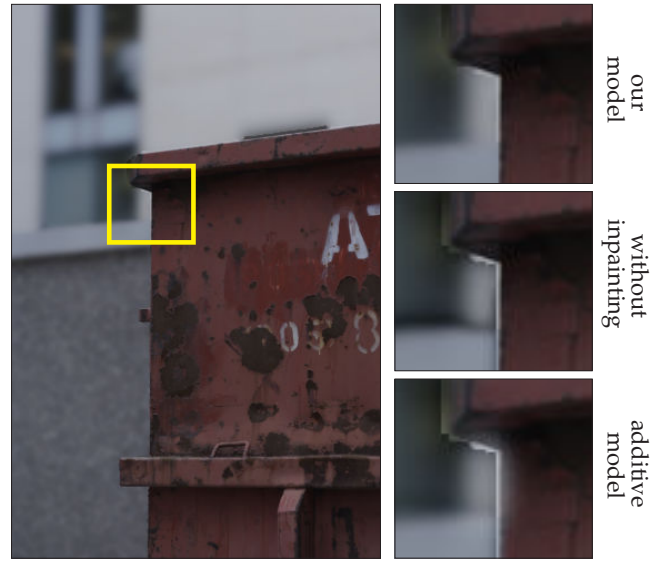

Fig. 9. Layered image formation results at occlusion boundaries. *Left:* Tone-mapped HDR image of the DUMP-STER dataset, for an extrapolated aperture (f/1). *Top inset:* Our model handles occlusions in a visually realistic way. *Middle:* Without inpainting, *i.e.*, assuming zero radiance in occluded regions, the resulting darkening emphasizes pixels whose layer assignment has been misestimated, that are not otherwise noticeable. *Bottom:* An additive image formation model [12], [14] exhibits similar artifacts, plus erroneous spill from the occluded background layer.

or 705×469 pixels.[5]

Our image restoration algorithm follows the description in Sec. 6, alternating between 10 conjugate gradient steps each of image restoration and blur refinement, until convergence. We periodically apply the layer reordering and refinement procedure as well, both immediately after initialization and every 80 such steps. As Fig. 8 shows, the image restoration typically converges within the first 100 iterations, and beyond the first application, layer reordering and refinement has

---

5. See `http://www.cs.toronto.edu/~hasinoff/aperture/` for additional results and videos.

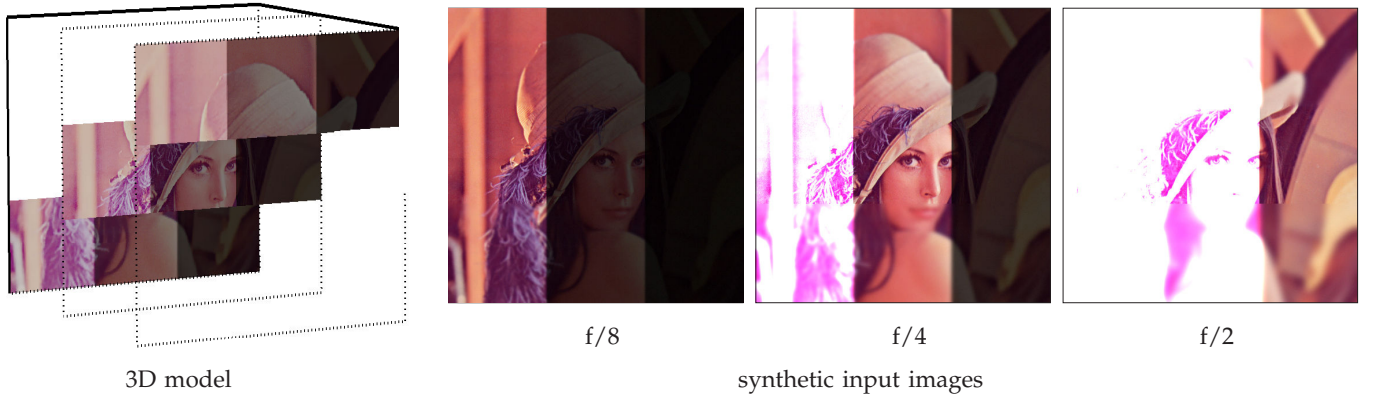| 3D model	synthetic input images
	f/8　　　　f/4　　　　f/2

Fig. 10. Synthetic LENA dataset. *Left:* Underlying 3D scene model, created from an HDR version of the Lena image. *Right:* Input images from applying our image formation model to the known 3D model, focused on the middle layer.

little effect. For all experiments we set the smoothing parameter to $\lambda = 0.002$.

Resynthesis with new camera settings. Upon completion of the image restoration stage, *i.e.*, once $(\mathbf{L}, \mathbf{A}, \sigma)$ has been estimated, we can apply the forward image formation model with arbitrary camera settings. This enables resynthesis of new images at near-interactive rates (Figs. 1,9–17).[6] Note that since we do not record the focus setting f at capture time, we fix the in-focus depth arbitrarily (*e.g.*, to $1.0\,\mathrm{m}$), which allows us to specify the depth of each layer in relative terms (*e.g.*, see Fig. 17). To synthesize photos with modified focus settings, we express the depth of the new focus setting as a fraction of the in-focus depth.[7]

Note that while camera settings can also be extrapolated, this functionality is somewhat limited. In particular, while extrapolating larger apertures than the maximum attainable by the lens lets us model exposure changes and increased defocus for each depth layer (Fig. 9), the depth resolution of our layered model is limited by the maximum lens aperture [30].

To demonstrate the benefit of our layered occlusion model for resynthesis, we compared our resynthesis results at layer boundaries with those obtained using alternative methods. As shown in Fig. 9, our layered occlusion model produces visually realistic output, even in the absence of pixel-accurate layer assignment. Our model is a significant improvement over the typical additive model of defocus [12], [14], which shows objectionable rendering artifacts at layer boundaries. Importantly, our layered occlusion model is accurate enough that we can resolve the correct layer ordering in

6. In order to visualize the exposure range of the recovered HDR radiance, we apply tone-mapping using a simple global operator of the form $T(x) = \frac{x}{1+x}$.

7. For ease of comparison, when changing the focus setting synthetically, we do not resynthesize geometric distortions such as image magnification. Similarly, we do not simulate the residual radiometric distortions $\hat{\mathbf{R}}$, such as vignetting. All these lens-specific artifacts can be simulated if desired.

all our experiments (except for one error in the DOORS dataset), simply by applying brute force search and testing which ordering leads to the smallest objective.

Synthetic data: LENA dataset. To enable comparison with ground truth, we tested our approach using a synthetic dataset (Fig. 10). This dataset consists of an HDR version of the 512×512 pixel Lena image, where we simulate HDR by dividing the image into three vertical bands and artificially exposing each band. We decomposed the image into layers by assigning different depths to each of three horizontal bands, and generated the input images by applying the forward image formation model, focused on the middle layer. Finally, we added Gaussian noise to the input with a standard deviation of $1\,\%$ of the intensity range.

As Fig. 11 shows, the restoration and resynthesis agree well with the ground truth, and show no visually objectionable artifacts, even at layer boundaries. The results show denoising throughout the image and even demonstrate good performance in regions that are both dark and defocused. Such regions constitute a worst case for our method, since they are dominated by noise for narrow apertures and are strongly defocused for wide apertures. Despite the challenge presented by these regions, our image restoration framework handles them naturally, because our formulation with TV regularization encourages the "deconvolution" of blurred intensity edges while simultaneously suppressing noise (Fig. 12a, inset). In general, however, weaker high-frequency detail cannot be recovered from strongly-defocused regions.

We also used this dataset to test the effect of using different numbers of input images spanning the same range of apertures from f/8 to f/2 (Table 1). As Fig. 13 shows, using only 2 input images significantly deteriorates the restoration results. As expected, using more input images improves the restoration, particularly with respect to recovering detail in dark and defocused regions, which benefit from the noise reduction that comes from additional images.
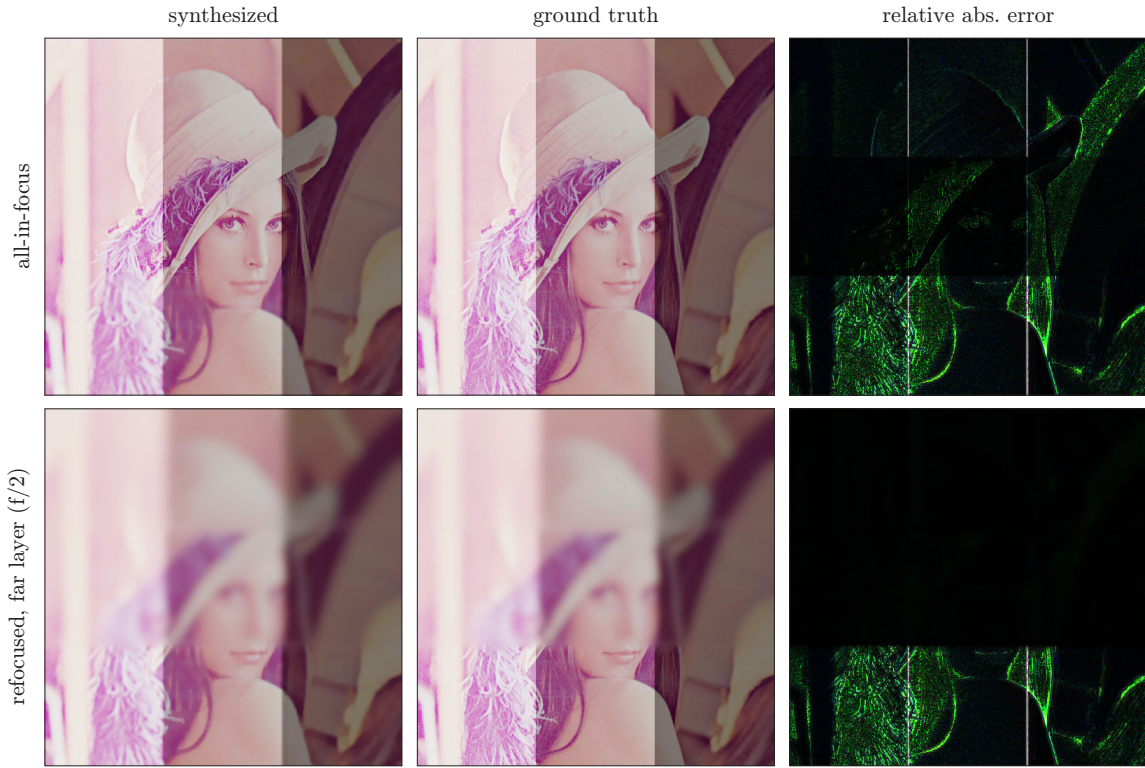
Fig. 11. Resynthesis results for the LENA dataset, shown tone-mapped, agree visually with ground truth. Note the successful smoothing and sharpening . The remaining errors are mainly due to the loss of the highest frequency detail caused by our image restoration and denoising. Because of the high dynamic range, we visualize the error in relative terms, as a fraction of the ground truth radiance.
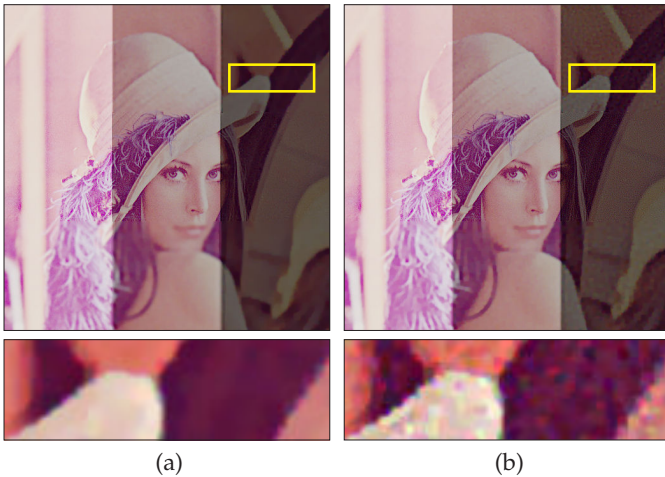

(a)　　　　　　　　(b)

Fig. 12. Effect of TV weighting. We show the all-in-focus HDR restoration result for the LENA dataset, tone-mapped and with enhanced contrast for the inset: (a) weighting the TV penalty according to effective exposure using Eq. (12), and (b) without weighting. In the absence of TV weighting, dark scene regions give rise to little TV penalty, and therefore get relatively under-smoothed. In both cases, TV regularization shows characteristic blocking into piecewise smooth regions.

TABLE 1
Restoration error for the LENA dataset, using different numbers of input images spanning the aperture range f/8–f/2. All errors are measured with respect to the ground truth HDR all-in-focus radiance.

num. input images	f-stops apart	RMS error	RMS rel. error	median rel. error
2	4	0.0753	13.2 %	2.88 %
3	2	0.0737	11.7 %	2.27 %
5	1	0.0727	11.4 %	1.97 %
9	1/2	0.0707	10.8 %	1.78 %
13	1/3	0.0688	10.6 %	1.84 %

DUMPSTER dataset. This outdoor scene has served as a running example throughout the article (Figs. 1, 5-9). It is composed of three distinct and roughly fronto-parallel layers: a background building, a pebbled wall, and a rusty dumpster. The foreground dumpster is darker than the rest of the scene and is almost in-focus. Although the layering recovered by the restoration is not pixel-accurate at the boundaries, resynthesis with new camera settings yields visually realistic results (Figs. 1 and 9).

PORTRAIT dataset. This portrait was captured indoors in a dark room, using only available light from the background window (Fig. 14). The subject is nearly in focus and is very dark compared to the background
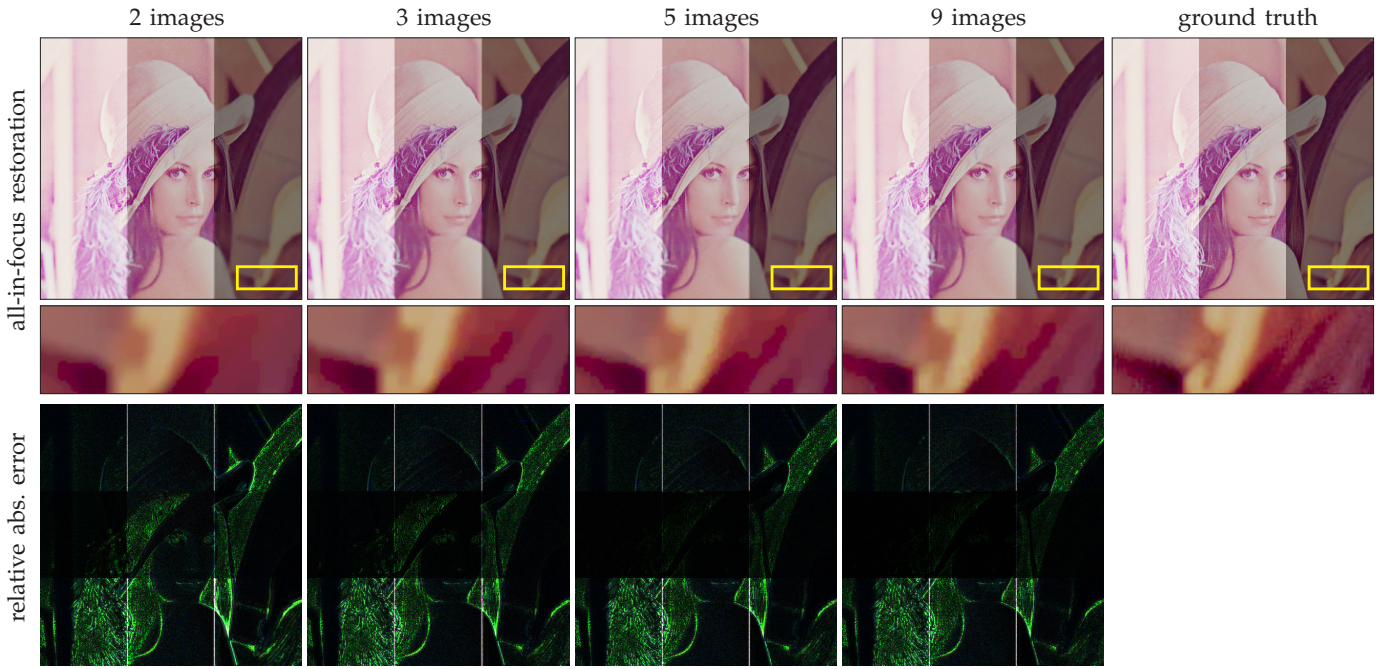
Fig. 13. Effect of the number of input images for the LENA dataset. *Top of row:* Tone-mapped all-in-focus HDR restoration. For better visualization, the inset is shown with enhanced contrast. *Bottom of row:* Relative absolute error, compared to the ground truth in-focus HDR radiance.

buildings outside; an even darker chair sits defocused in the foreground. Note that while the final layer assignment is only roughly accurate (*e.g.*, near the subject's right shoulder), the discrepancies are restricted mainly to low-texture regions near layer boundaries, where layer membership is ambiguous and has little influence on resynthesis. In this sense, our method is similar to image-based rendering from stereo [31], [32] where reconstruction results that deviate from ground truth in "unimportant" ways can still lead to visually realistic new images. Slight artifacts can be observed at the boundary of the chair, in the form of an over-sharpened dark stripe running along its arm. This part of the scene was under-exposed even in the widest-aperture image, and the blur diameter was apparently estimated too high, perhaps due to over-fitting the background pixels that were incorrectly assigned to the chair.

DOORS dataset. This architectural scene was captured outdoors at twilight and consists of a sloping wall containing a row of rusty doors, with a more brightly illuminated background (Fig. 15). The sloping, hallway-like geometry constitutes a challenging test for our method's ability to handle scenes that violate our piecewise fronto-parallel scene model. As the results show, despite the fact that our method decomposes the scene into six fronto-parallel layers, the recovered layer ordering is almost correct, and our restoration allows us to resynthesize visually realistic new images. Note that the reduced detail for the tree in the background is due to scene motion caused by wind over the $1\,\mathrm{s}$ total

capture time.

Failure case: MACRO dataset. Our final sequence was a macro still life scene, captured using a $10\,\mathrm{mm}$ extension tube to reduce the minimum focusing distance of the lens, and to increase the magnification to approximately life size (1:1). The scene is composed of a miniature glass bottle whose inner surface is painted, and a dried bundle of green tea leaves (Fig. 16). This is a challenging dataset for several reasons: the level of defocus is severe outside the very narrow depth of field, the scene consists of both smooth and intricate geometry (bottle and tea leaves, respectively), and the reflections on the glass surface only become focused at incorrect virtual depths. The initial segmentation leads to a very coarse decomposition into layers that is not improved by our optimization. While the resynthesis results for this scene suffer from strong artifacts, the gross structure, blur levels, and ordering of the scene layers are still recovered correctly. The worst artifacts are the bright "cracks" occurring at layer boundaries. These are due to a combination of incorrect layer segmentation and our diffusion-based inpainting method.

A current limitation of our method is that our scheme for re-estimating the layering is not always effective. Although pixels not reproducing the input images sometimes indicate incorrect layer labels, they may also indicate overfitting and other sources of error such as imperfect calibration. Fortunately, even when the layering is not estimated exactly, our layered occlusion model often leads to visually realistic resynthesized images (*e.g.*, Figs. 9 and 14).
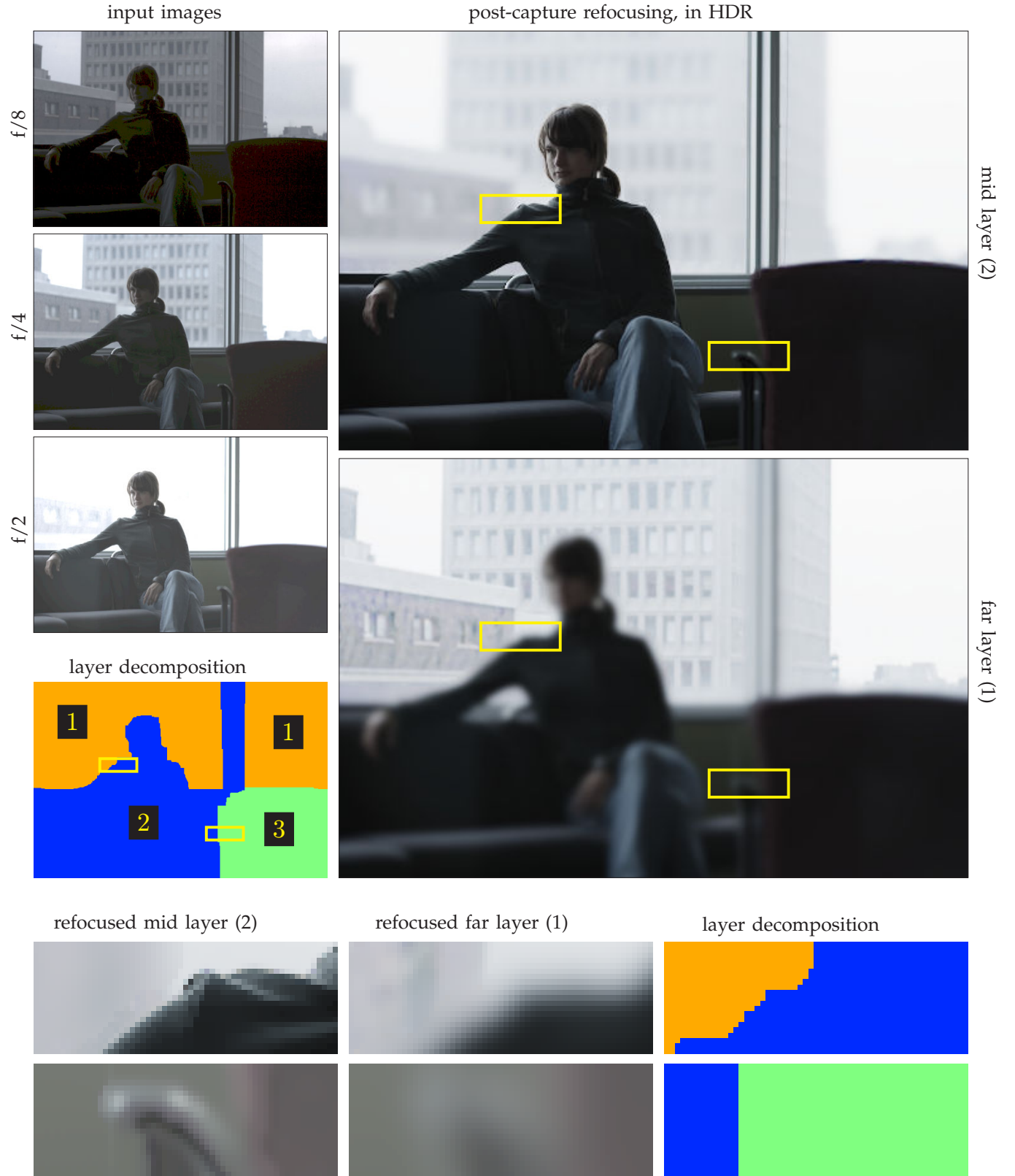
Fig. 14. PORTRAIT dataset. The input images are visualized with strong gamma correction ($\gamma = 3$) to display the high dynamic range of the scene, and show significant posterization artifacts. Although the final layer assignment has errors in low-texture regions near layer boundaries, the restoration results are sufficiently accurate to resynthesize visually realistic new images. We demonstrate refocusing in HDR with tone-mapping, simulating the widest input aperture (f/2).
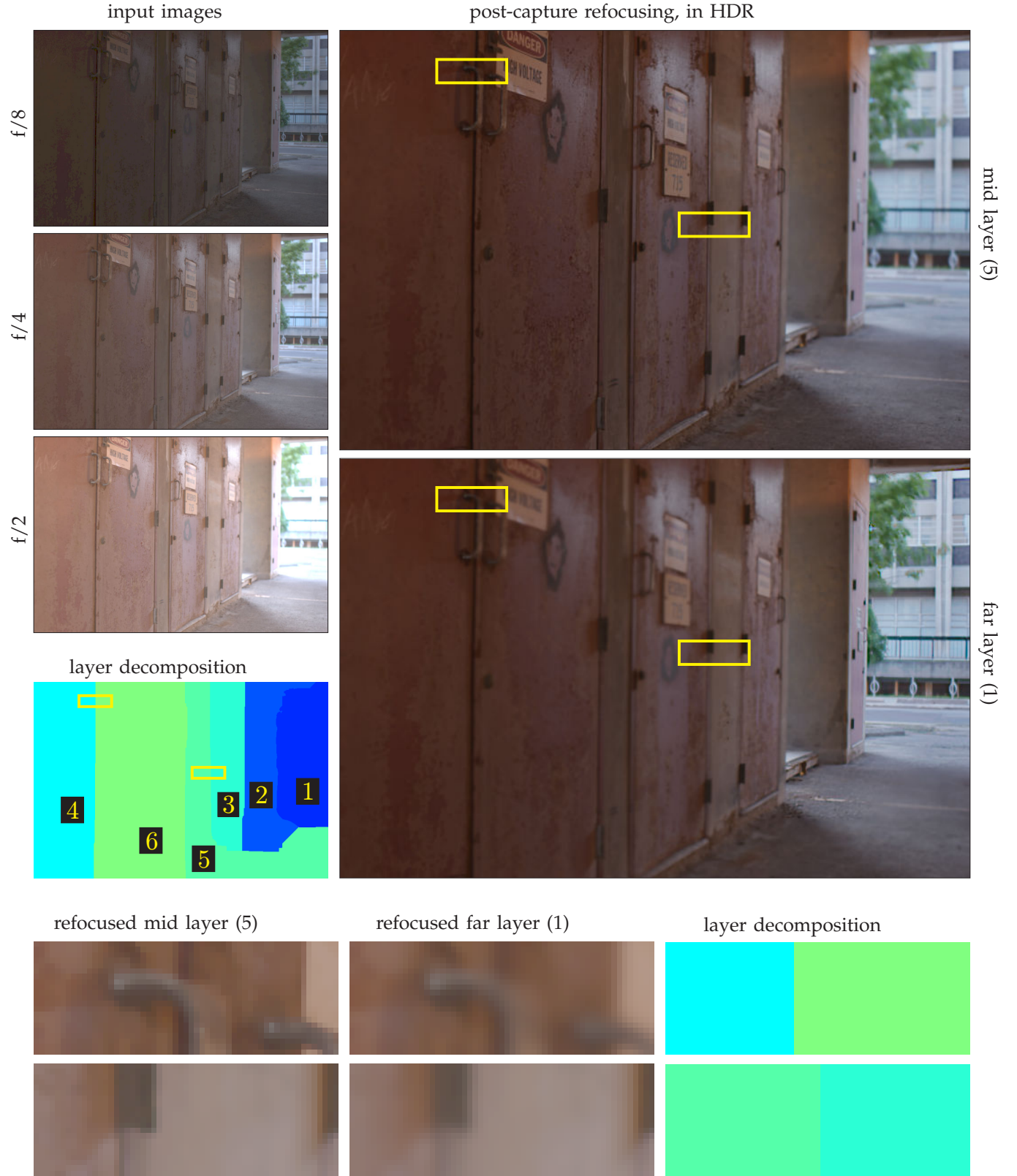
Fig. 15. DOORS dataset. The input images are visualized with strong gamma correction ($\gamma = 3$) to display the high dynamic range of the scene. Our method approximates the sloping planar geometry of the scene using a small number of fronto-parallel layers. Despite this approximation, and an incorrect layer ordering estimated for the leftmost layer, our restoration results are able to resynthesize visually realistic new images. We demonstrate refocusing in HDR with tone-mapping, simulating the widest input aperture (f/2).
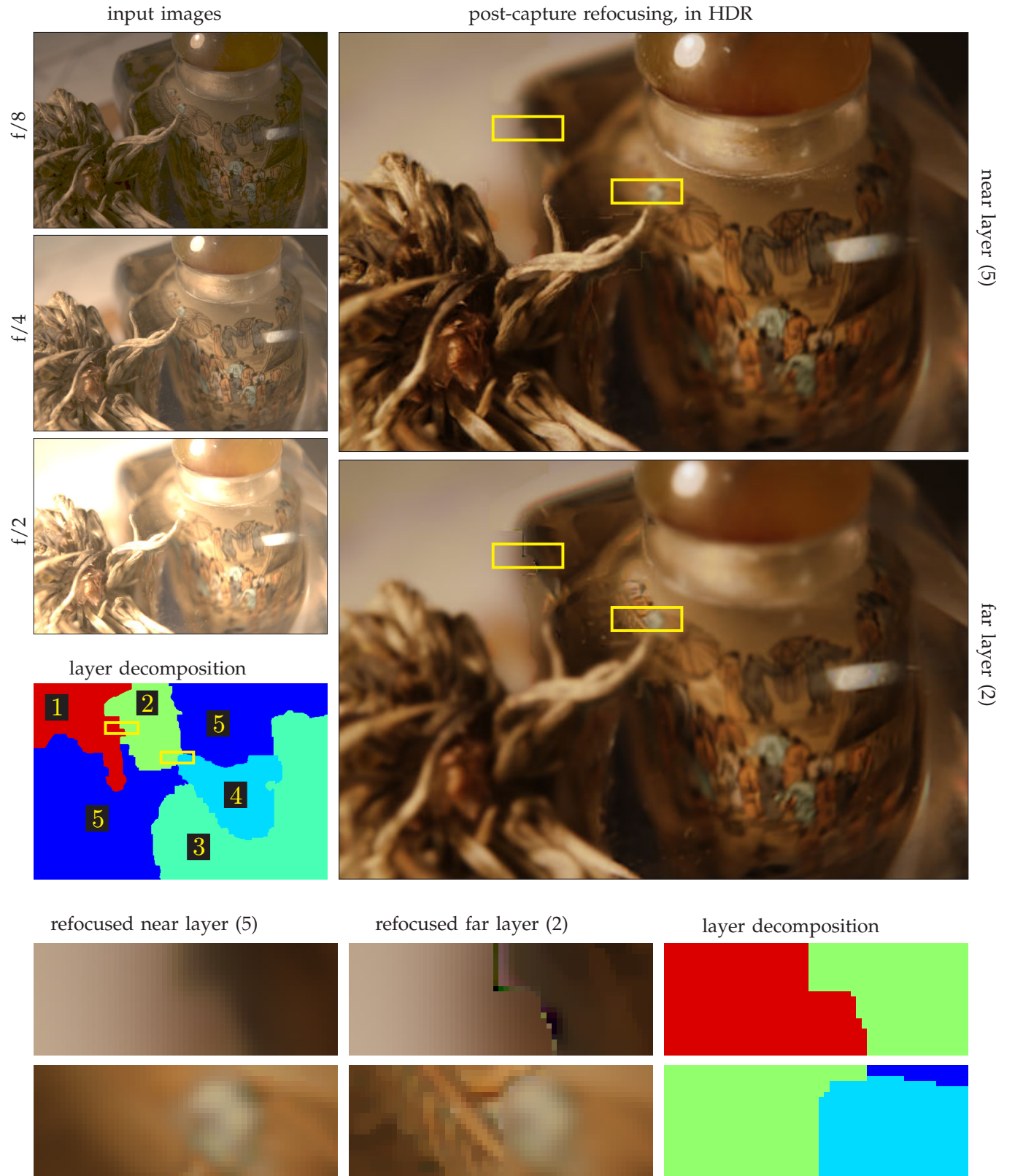
Fig. 16. MACRO dataset (failure case). The input images are visualized with strong gamma correction ($\gamma = 3$) to display the high dynamic range of the scene. The recovered layer segmentation is very coarse, and significant artifacts are visible at layer boundaries, due to a combination of the incorrect layer segmentation and our diffusion-based inpainting. We demonstrate refocusing in HDR with tone-mapping, simulating the widest input aperture (f/2).
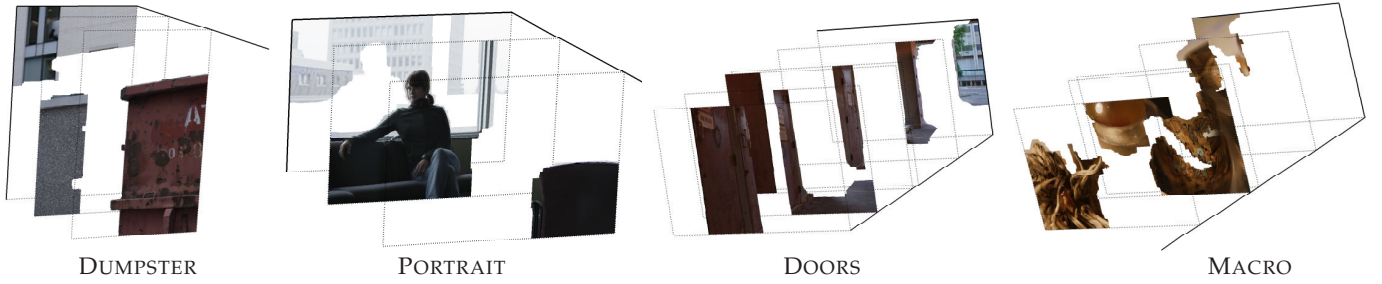
| DUMPSTER | PORTRAIT | DOORS | MACRO |

Fig. 17. Gallery of restoration results for the real datasets. We visualize the recovered layers in 3D using the relative depths defined by their blur diameters and ordering.

9 CONCLUDING REMARKS

We showed that multiple-aperture photography leads to a unified restoration framework for decoupling the effects of defocus and exposure, permitting HDR photography and post-capture control of a photo's camera settings. From a user interaction perspective, one can imagine creating new controls to navigate the space of camera settings offered by our representation. In fact, our recovered scene model is rich enough to synthesize arbitrary per-layer defocus and to enable special effects such as compositing new objects into the scene.

For future work, we are interested in addressing motion between exposures, caused by hand-held photography or subject motion. Although we have experimented with simple image registration methods, it would be beneficial to integrate a layer-based parametric model of optical flow directly into the overall optimization. We are also interested in improving the efficiency of our technique by exploring multi-resolution variants of the basic method.

While each layer is currently modeled as a binary mask, it is possible to represent each layer with fractional alpha values, in order to improve resynthesis at boundary pixels that are mixtures of background and foreground. Our image formation model (Sec. 4) already handles layers with general alpha mattes, and it should be straightforward to process our layer estimates in the vicinity of the initial hard boundaries using existing matting techniques [31], [33]. This color-based matting may also be useful help refine the initial layering we estimate using depth-from-defocus.

APPENDIX A
ANALYTIC GRADIENTS FOR LAYER-BASED RESTORATION

Because our image formation model is a composition of linear operators plus clipping, the gradients of the objective function defined in Eqs. (8)–(10) have a compact analytic form.

Intuitively, our image formation model can be thought of as spatially-varying linear filtering, analogous to convolution ("distributing" image intensity according to the blur diameters and layering). Thus, the adjoint operator that defines its gradients corresponds to spatially-varying linear filtering as well, analogous to correlation ("gathering" image intensity) [34].

Simplified gradient formulas. For clarity, we first present gradients of the objective function assuming a single aperture, a, without inpainting:

$$\frac{\partial \mathcal{O}}{\partial \mathbf{L}} = e_a \mathbf{U}_a \sum_{k=1}^{K} [\mathbf{A}_k \mathbf{M}_k \boldsymbol{\Delta} \star B_{\sigma_k}] + \frac{\partial \|\mathbf{L}\|_\beta}{\partial \mathbf{L}} \quad (14)$$

$$\frac{\partial \mathcal{O}}{\partial \sigma_k} = e_a \mathbf{U}_a \sum_{x,y} \left[\sum_{j=1}^{K} \left[\mathbf{A}_j \mathbf{M}_j \boldsymbol{\Delta} \star \frac{\partial B_{\sigma_j}}{\partial \sigma_j} \right] \right] \mathbf{A}_k \mathbf{L} \ , \quad (15)$$

where $\star$ denotes 2D correlation, and the binary mask

$$\mathbf{U}_a = \left[e_a \overline{\mathbf{L}} < 1 \right] \quad (16)$$

indicates which pixels in the synthesized input image are unsaturated, thereby assigning zero gradients to over-saturated pixels. This definition resolves the special case $e_a \overline{\mathbf{L}} = 1$, at which point the gradient of Eq. (10) is discontinuous. Since all matrix multiplications above are pixel-wise, we have omitted the operator $\cdot$ for brevity.

The only expression left to specify is the gradient for the regularization term in Eq. (12):

$$\frac{\partial \|\mathbf{L}\|_\beta}{\partial \mathbf{L}} = -\operatorname{div}\left(\frac{w(\mathbf{L})^2 \nabla \mathbf{L}}{\sqrt{\left(w(\mathbf{L}) \|\nabla \mathbf{L}\|\right)^2 + \beta}} \right) \ , \quad (17)$$

where div is the divergence operator. This formula is a slight generalization of a previous treatment for the total variation norm [23], but it incorporates per-pixel weights, $w(\mathbf{L})$, to account for high dynamic range.

Multiple aperture settings. The generalization to the multiple aperture settings is straightforward. We add an outer summation over aperture, and relate blur diameter across aperture using scale factors that follow from Eq. (3), $s_a = \frac{D_a}{D_A}$. See footnote 3 (p. 4) for more detail about how we compute these scale factors in practice.

Inpainting. To generalize the gradient formulas to include inpainting, we assume that the inpainting operator

$$\frac{\partial \mathcal{O}}{\partial \mathbf{L}} \quad = \quad \sum_{a=1}^{A} e_a \mathbf{U}_a \left[\sum_{k=1}^{K} \mathcal{I}_k^{\dagger} \left[\mathbf{A}'_k \mathbf{M}_k \boldsymbol{\Delta}_a \star B_{(s_a \sigma_k)} \right] \right] \; + \; \frac{\partial \|\mathbf{L}\|_{\beta}}{\partial \mathbf{L}} \tag{18}$$

$$\frac{\partial \mathcal{O}}{\partial \sigma_k} \quad = \quad \sum_{a=1}^{A} s_a e_a \mathbf{U}_a \left[\sum_{x,y} \left[\sum_{j=1}^{K} \mathcal{I}_k^{\dagger} \left[\mathbf{A}'_j \mathbf{M}_j \boldsymbol{\Delta}_a \star \frac{\partial B_{(s_a \sigma_j)}}{\partial (s_a \sigma_j)} \right] \right] \mathbf{A}'_k \mathbf{L} \right] \quad . \tag{19}$$

for each layer k,

$$\mathcal{I}_k[\mathbf{L}] = \mathbf{A}'_k \mathbf{L} + \mathbf{A}''_k \mathbf{L}''_k \quad , \tag{20}$$

can be expressed as a linear function of radiance. This model covers many existing inpainting methods, including choosing the nearest unoccluded pixel, PDE-based diffusion [27], and exemplar-based inpainting.

To compute the gradient, we need to determine the adjoint of the inpainting operator, $\mathcal{I}_k^{\dagger}[\cdot]$, which has the effect of "gathering" the inpainted radiance from its occluded destination and "returning" it to its unoccluded source. In matrix terms, if the inpainting operator is written as a large matrix left-multiplying the flattened scene radiance, $\mathcal{I}_k$, the adjoint operator is simply its transpose, $\mathcal{I}_k^T$.

Gradient formulas. Putting everything together, we obtain the final gradients in Eqs. (18)–(19).

ACKNOWLEDGMENTS

REFERENCES

[1] T. Mitsunaga and S. K. Nayar, "Radiometric self calibration," in *Proc. Computer Vision and Pattern Recognition*, 1999, pp. 1374–1380.

[2] E. Eisemann and F. Durand, "Flash photography enhancement via intrinsic relighting," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 673–678, 2004.

[3] A. Agarwala, M. Dontcheva, M. Agrawala, S. Drucker, A. Colburn, B. Curless, D. Salesin, and M. Cohen, "Interactive digital photomontage," in *Proc. ACM SIGGRAPH*, 2004, pp. 294–302.

[4] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," Dept. Computer Science, Stanford University, Tech. Rep. CTSR 2005-02, 2005.

[5] A. Isaksen, L. McMillan, and S. J. Gortler, "Dynamically reparameterized light fields," in *Proc. ACM SIGGRAPH*, 2000, pp. 297–306.

[6] Flickr HDR group, http://www.flickr.com/groups/hdr/.

[7] N. Asada, H. Fujiwara, and T. Matsuyama, "Seeing behind the scene: Analysis of photometric properties of occluding edges by the reversed projection blurring model," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 20, no. 2, pp. 155–167, 1998.

[8] P. Favaro and S. Soatto, "Seeing beyond occlusions (and other marvels of a finite lens aperture)," in *Proc. Computer Vision and Pattern Recognition*, vol. 2, 2003, pp. 579–586.

[9] S. S. Bhasin and S. Chaudhuri, "Depth from defocus in presence of partial self occlusion," in *Proc. International Conference on Computer Vision*, vol. 2, 2001, pp. 488–493.

[10] M. McGuire, W. Matusik, H. Pfister, J. F. Hughes, and F. Durand, "Defocus video matting," in *Proc. ACM SIGGRAPH*, 2005, pp. 567–576.

[11] A. Smith and J. Blinn, "Blue screen matting," in *Proc. ACM SIGGRAPH*, 1996, pp. 259–268.

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

[13] H. Jin and P. Favaro, "A variational approach to shape from defocus," in *Proc. European Conference on Computer Vision*, vol. 2, 2002, pp. 18–30.

[14] M. Šorel and J. Flusser, "Simultaneous recovery of scene structure and blind restoration of defocused images," in *Proc. Computer Vision Winter Workshop*, 2006, pp. 40–45.

[15] K. Aizawa, K. Kodama, and A. Kubota, "Producing object-based special effects by fusing multiple differently focused images," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 10, no. 2, pp. 323–330, Mar. 2000.

[16] S. Chaudhuri, "Defocus morphing in real aperture images," *J. Optical Society of America A*, vol. 22, no. 11, pp. 2357–2365, Nov. 2005.

[17] D. B. Goldman, "Vignette and exposure calibration and compensation," in *Proc. International Conference on Computer Vision*, 2005, pp. 899–906.

[18] M. Aggarwal and N. Ahuja, "A pupil-centric model of image formation," *International Journal of Computer Vision*, vol. 48, no. 3, pp. 195–214, 2002.

[19] A. P. Pentland, "A new sense for depth of field," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 9, no. 4, pp. 523–531, Jul. 1987.

[20] S. W. Hasinoff and K. N. Kutulakos, "Confocal stereo," in *Proc. European Conference on Computer Vision*, vol. 1, 2006, pp. 620–634.

[21] L. Zhang and S. K. Nayar, "Projection defocus analysis for scene capture and image display," in *Proc. ACM SIGGRAPH*, 2006, pp. 907–915.

[22] Y. Y. Schechner and S. K. Nayar, "Generalized mosaicing: High dynamic range in a wide field of view," *International Journal of Computer Vision*, vol. 53, no. 3, pp. 245–267, 2004.

[23] C. Vogel and M. Oman, "Fast, robust total variation based reconstruction of noisy, blurred images," *IEEE Trans. on Image Processing*, vol. 7, no. 6, pp. 813–824, Jun. 1998.

[24] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 26, no. 9, pp. 1124–1137, Sep. 2004.

[25] M. Subbarao and G. Surya, "Depth from defocus: A spatial domain approach," *International Journal of Computer Vision*, vol. 13, no. 3, pp. 271–294, Dec. 1994.

[26] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 23, no. 11, pp. 1222–1239, Nov. 2001.

[27] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," in *Proc. ACM SIGGRAPH*, 2000, pp. 417–424.

[28] A. Levin, R. Fergus, F. Durand, and W. T. Freeman, "Image and depth from a conventional camera with a coded aperture," in *Proc. ACM SIGGRAPH*, 2007.

[29] F. Moreno-Noguer, P. N. Belhumeur, and S. K. Nayar, "Active refocusing of images and videos," in *Proc. ACM SIGGRAPH*, 2007.

[30] Y. Y. Schechner and N. Kiryati, "Depth from defocus vs. stereo: How different really are they?" *International Journal of Computer Vision*, vol. 39, no. 2, pp. 141–162, Sep. 2000.

[31] C. L. Zitnick and S. B. Kang, "Stereo for image-based rendering using image over-segmentation," *International Journal of Computer Vision*, vol. 75, no. 1, pp. 49–65, 2007.

[32] A. Fitzgibbon, Y. Wexler, and A. Zisserman, "Image-based rendering using image-based priors," *International Journal of Computer Vision*, vol. 63, no. 2, pp. 141–151, Jul. 2005.

[33] S. W. Hasinoff, S. B. Kang, and R. Szeliski, "Boundary matting for view synthesis," *Computer Vision and Image Understanding*, vol. 103, no. 1, pp. 22–32, Jul. 2006. [Online]. Available: http://dx.doi.org/10.1016/j.cviu.2006.02.005

[34] M. Šorel, "Multichannel blind restoration of images with space-variant degradations," Ph.D. dissertation, Charles University in Prague, Dept. of Software Engineering, 2007.

Samuel W. Hasinoff received the BS degree in computer science from the University of British Columbia in 2000, and the MS and PhD degrees in computer science from the University of Toronto in 2002 and 2008, respectively. He is currently an NSERC Postdoctoral Fellow at the Massachusetts Institute of Technology. His research interests include computer vision and computer graphics, with a current focus on computational photography. In 2006, he received an honorable mention for the Longuet-Higgins Best Paper Award at the European Conference on Computer Vision. He is a member of the IEEE.

Kiriakos N. Kutulakos received the BA degree in computer science at the University of Crete, Greece in 1988, and the MS and PhD degrees in computer science from the University of Wisconsin, Madison in 1990 and 1994, respectively. Following his dissertation work, he joined the University of Rochester where he was an NSF Postdoctoral Fellow and later an assistant professor until 2001. He is currently an associate professor of computer science at the University of Toronto. He won the Best Student Paper Award at CVPR'94, the Marr Prize in 1999, a Marr Prize Honorable Mention in 2005 and a Best Paper Honorable Mention at ECCV'06. He is the recipient of a CAREER award from the US National Science Foundation, a Premier's Research Excellence Award from the government of Ontario, and an Alfred P. Sloan Research Fellowship. He served as program co-chair of CVPR 2003 and is currently an associate editor of the IEEE Transactions on Pattern Analysis and Machine Intelligence. He is a member of the IEEE.