Removing Image Artifacts Due to Dirty Camera Lenses and Thin Occluders

Jinwei Gu
Columbia University

Ravi Ramamoorthi
University of California at Berkeley

Peter Belhumeur
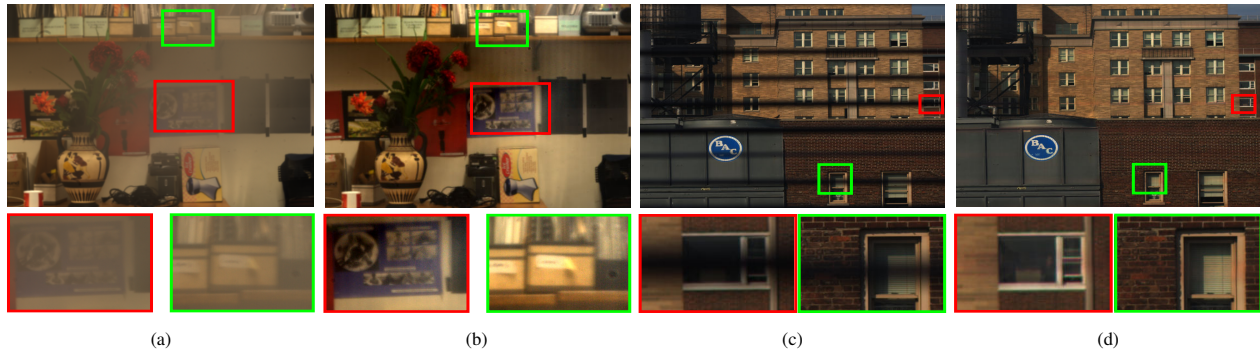Columbia University

Shree Nayar
Columbia University

Figure 1: *Removal of image artifacts due to dirty camera lenses and thin occluders. (a) An image taken with a Canon EOS 20D camera equipped with a dirty lens, showing significant artifacts due to attenuation and scattering of lens dirt. (b) The recovered image with the artifacts removed. (c) A photograph taken inside a room through a window shutter exhibits black stripe artifacts due to occlusion. (d) By taking pictures with different apertures, we can effectively remove the artifacts. Details are shown in the insets.*

Abstract

Dirt on camera lenses, and occlusions from thin objects such as fences, are two important types of artifacts in digital imaging systems. These artifacts are not only an annoyance for photographers, but also a hindrance to computer vision and digital forensics. In this paper, we show that both effects can be described by a single image formation model, wherein an intermediate layer (of dust, dirt or thin occluders) both attenuates the incoming light and scatters stray light towards the camera. Because of camera defocus, these artifacts are low-frequency and either additive or multiplicative, which gives us the power to recover the original scene radiance pointwise. We develop a number of physics-based methods to remove these effects from digital photographs and videos. For dirty camera lenses, we propose two methods to estimate the attenuation and the scattering of the lens dirt and remove the artifacts – either by taking several pictures of a structured calibration pattern beforehand, or by leveraging natural image statistics for post-processing existing images. For artifacts from thin occluders, we propose a simple yet effective iterative method that recovers the original scene from multiple apertures. The method requires two images if the depths of the scene and the occluder layer are known, or three images if the depths are unknown. The effectiveness of our proposed methods are demonstrated by both simulated and real experimental results.

CR Categories: I.4.3 [Image Processing and Computer Vision]: Enhancement—Grayscale Manipulation

Keywords: image enhancement, computational photography

1 Introduction

A common assumption in computer graphics, as well as in digital photography and imaging systems, is that the radiance emitted from a scene point is observed directly at the sensor. However, there are often physical layers or media lying between the scene and the imaging system. For example, the lenses of consumer digital cameras, or the front windows of security cameras, often accumulate various types of contaminants over time (*e.g.*, fingerprints, dust, dirt). Artifacts from a dirty camera lens are shown in Fig. 1(a). Figure 1(c) shows another example of the undesired artifacts caused by a layer of thin occluders (*e.g.*, fences, meshes, window shutters, curtains, tree branches), where a photograph is taken inside a room through a window shutter that partially obstructs the scene. Both artifacts are annoying for photographers, and may also damage important scene information for applications in computer vision or digital forensics.

Of course, a simple solution is to clean the camera lens, or choose a better spot to retake pictures. However, this is impossible for existing images, and impractical for some applications like outdoor security cameras, underwater cameras or covert surveillance behind a fence. Therefore, we develop new ways to take the pictures, and new computational algorithms to remove dirty-lens and thin-occluder artifacts. Unlike image inpainting and hole-filling methods, our algorithms rely on an understanding of the physics of image formation to directly recover the image information in a pointwise fashion, given that each point is partially visible in at least one of the captured images.

Both artifacts can be described using a single image formation model as shown in Fig. 2, where an intermediate layer between the target scene and the camera lens affects the image irradiance in two ways: (1) **Attenuation** where the scene radiance is reduced, either by absorption (in the case of lens dirt) or obstruction (in the case of thin occluders); and (2) **Intensification** where the intermediate layer itself will contribute some radiance to the image sensor, either by scattering the light from other directions (*e.g.*, scattering of sunlight by lens dirt) or by reflecting the light from the surface of the layer. Intuitively, attenuation tends to make the affected regions darker while intensification tends to make the regions brighter.[1] Because of camera defocus, both attenuation and intensification are low-frequency, pointwise operations. In other words, *the high fre-*

---

[1]In the case of lens dirt, such effects have been reported previously in both optics [Willson et al. 2005] and computer graphics [Gu et al. 2007].

quencies in the original scene radiance will be partially preserved in the degraded images. This can be seen in the insets of Fig. 1 where the edges of the background are still partially visible in the degraded images. Based on these observations, we develop several fully automatic methods to estimate and remove these artifacts from photographs and videos:

Dirty Camera Lens: We demonstrate two methods to estimate the attenuation and the scattering patterns (Secs. 4.2 and 4.3). First, if we have access to the camera, the attenuation and scattering patterns of the lens dirt can be directly measured from multiple pictures (≥ 2) of a structured pattern. Second, for existing images or situations where we do not have access to the camera, we show that the attenuation and scattering patterns can be estimated from a collection of photographs taken with the same dirty lens camera by using prior knowledge on natural image statistics.

Once the attenuation and scattering patterns of the lens dirt are known, the artifacts are removed from individual photographs by enforcing sparsity in the recovered images' gradients. Figure 1(b) shows an example of the image recovered from Fig. 1(a).

Thin Occluders: For thin occluders, we consider a special case where the reflected light from the occluders is negligible compared to the scene radiance passing through. We develop an iterative method to remove the image artifacts from two input images with different apertures, if the depth of the occluder layer is known (Sec. 5.1). More generally, we show that by using three apertures, we can both remove the image artifacts and estimate the depth of the occluding layer (Sec. 5.2). Figure 1(d) shows an example of thin occluder removal, with the insets revealing more detailed information from the scene.

2 Related Work

Image Inpainting and Hole-Filling: A variety of inpainting and texture synthesis techniques are used to correct imperfections in photographs [Bertalmio et al. 2000; Efros and Freeman 2001; Sun et al. 2005]. An interesting recent work [Liu et al. 2008] uses symmetry to remove structured artifacts (*e.g.*, fences and meshes) from input images. These methods need no knowledge of physics, and rely purely on neighboring regions to *synthesize* information in the affected regions. In contrast, since the information of the original scene is still (partially) accessible in our case, we are able to take a physically-based approach and recover the original scene point-wise, which is expected to be more faithful to its actual structure. Note that our pointwise operations and image inpainting methods are not mutually exclusive – the former can be used where the scene is partially visible, after which the latter can be used on the remaining, completely-blocked areas.

Modeling and Removing Camera Artifacts: [Talvala et al. 2007] studied camera-veiling glare which is, in effect, a uniform loss of contrast in the images. [Raskar et al. 2008] proposed the use of a light field camera to remove lens glare. Recently, [Koreban and Schechner 2009] showed lens glare can be used to estimate the geometry of the source and the camera. Another body of work considered image artifacts caused by scattering of participating media and developed techniques for dehazing [Schechner et al. 2003], contrast restoration in bad weather [Narasimhan and Nayar 2003], and underwater imaging [Schechner and Karpel 2005]. Two interesting recent work on dehazing [Fattal 2008; He et al. 2009] combines both statistical and physically-based approaches. While we draw inspiration from these methods, and the broader area of computational photography, we focus on different visual effects.

In the context of dirty camera lenses, [Willson et al. 2005] considered the image appearance of dust that lies on a transparent lens cover. They reported both "dark dust artifacts" (*i.e.*, attenuation)
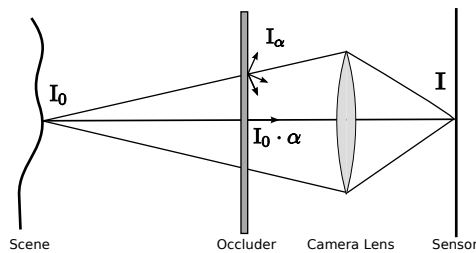

Figure 2: *Image formation model: $\boldsymbol{\alpha} \in [0, 1]$ is the attenuation pattern of the intermediate layer, i.e., the fraction of light transmitted (0 = completely blocked). The camera is focused on the target scene. The final image $\mathbf{I}$ is the sum of the attenuated light from the background after the defocus blur, $\mathbf{I}_0 \cdot (\boldsymbol{\alpha} * \mathbf{k})$, and the light emitted from the intermediate layer itself, $\mathbf{I}_{\boldsymbol{\alpha}} * \mathbf{k}$.*

and "bright dust artifacts" (*i.e.*, intensification) but only handled the "dark dust artifacts." [Zhou and Lin 2007] studied artifacts caused by the attenuation due to dust on image sensors. In contrast, we consider both the attenuation and intensification effects caused by contaminants on camera lenses.

Depth from Defocus and Large Apertures: To remove the artifacts introduced by a layer of thin occluders, our methods rely on the defocus blur of the occluder layer, caused by the finite size of the aperture. In the limiting case of a very large aperture and the camera focused on the target scene, the occluder layer will be heavily defocused and completely blurred out in the final image, as demonstrated by recent work on synthetic apertures using dense camera arrays [Vaish et al. 2006]. For conventional consumer cameras, however, it is less practical to have such large apertures. Thus, this occlusion effect has almost always been considered a problem in depth from defocus methods [Watanabe and Nayar 1996].

Previous works [Favaro and Soatto 2003; Hasinoff and Kutulakos 2007] took occlusion into account and proposed methods to recover depth and scene textures from a stack of images with different focus or aperture settings. [McCloskey et al. 2007] recently proposed an interesting method to remove partial occlusion blur from a single image using geometric flow, although it is not obvious how to extend the method for multiple, complex occlusion objects such as curtains or tree branches. Occlusion has also been utilized for video matting [Durand et al. 2005], where the goal is to recover an accurate trimap on the boundary, rather than recovering the occluded scene. In contrast, our goal is primarily to recover the scene radiance, rather than depth or alpha matte. Thus, we require two (if the depth of the occluder layer is known), or three (if the depth of the occluder layer is unknown) images, which makes our method simpler and faster than previous works.

3 Image Formation Model

We first explain the image formation model used in this paper, distinguished by an intermediate layer between the camera and the target scene (either contaminants on camera lenses or thin occluders such as fences). We assume the target scene is always in focus in both cases. As shown in Fig. 2, the final image $\mathbf{I}(x, y)$ captured by the camera consists of two components. The first is **attenuation** where the radiance emitted from the target scene is attenuated by the intermediate layer. The second is **intensification**, where the intermediate layer itself contributes some radiance to the camera, by either scattering light from other directions in the environment or reflecting light from its surface. Suppose $\mathbf{I}_0(x, y)$ is the radiance of the target scene, $\boldsymbol{\alpha}(x, y) \in [0, 1]$ is the attenuation pattern of the intermediate layer, *i.e.*, the fraction of light transmitted (0 = completely blocked), and $\mathbf{I}_{\boldsymbol{\alpha}}(x, y)$ is the intensification term (*i.e.*, the

extra radiance from the intermediate layer itself). We have:

$$\mathbf{I} = \mathbf{I}_0 \cdot (\boldsymbol{\alpha} * \mathbf{k}) + \mathbf{I}_{\boldsymbol{\alpha}} * \mathbf{k}, \qquad (1)$$

where $\mathbf{k}(x, y)$ is the defocus blur kernel for the intermediate layer, and $*$ denotes image convolution.[2] There are several important features of Equation (1):

Pointwise multiplication preserves high frequencies: The term $\mathbf{I}_0 \cdot (\boldsymbol{\alpha} * \mathbf{k})$ involves a pointwise multiplication of the artifact-free image I_0 and the attenuation $\boldsymbol{\alpha} * \mathbf{k}$. This means that as long as $\boldsymbol{\alpha} * \mathbf{k} > 0$, all the high-frequency components in $\mathbf{I}_0$ will still be partially preserved in the final artifacted image $\mathbf{I}$. It is this property which lets us recover the original scene in a pointwise manner.

On the other hand, this property also shows a major limitation of our methods: in regions where $\boldsymbol{\alpha} * \mathbf{k} = 0$, no information on the original scene is captured in $\mathbf{I}$, and we have to rely on neighboring pixels to recover the original scene $\mathbf{I}_0$. Fortunately, this is usually not a problem for dirty-lens artifacts – the defocus blur for lens dirt is so large that the artifacts are always presented as a low frequency pattern in images, as shown in Fig.1(a). For thin-occluder artifacts such as fences or window shutters, the aperture size needs to be large enough to partially capture the background radiance in at least one of the captured images, as discussed in Sec. 6.

Intensification: The second term in Equation (1) corresponds to the intensification $\mathbf{I}_{\boldsymbol{\alpha}}$. It can come from multiple sources – for lens dirt, it is the scattering of the environmental lighting; for thin occluders, it is the appearance of the occluders. In both cases, we know that wherever the layer does not block any light (*i.e.*, $\boldsymbol{\alpha} = 1$), there will be no intensification either (*i.e.*, $\mathbf{I}_{\boldsymbol{\alpha}} = 0$).

Solving for I_0: Equation (1) shows that recovering the artifact-free image $\mathbf{I}_0$ from a single input image is ill-posed in general, because both $\boldsymbol{\alpha}$ and $\mathbf{I}_{\boldsymbol{\alpha}}$ are unknown. The defocus blur kernel $\mathbf{k}$ might also be unknown, since it depends on the depths of the scene and the intermediate layer. In the rest of the paper, we show how we can estimate some of these terms (either with a simple calibration step or by using multiple images) and then remove the artifacts.

4 Artifacts Caused by Dirty Camera Lenses

In the case of lens dirt, the attenuation $\boldsymbol{\alpha}$ is caused by the absorption of the contaminants, which can be modeled as $\boldsymbol{\alpha} \approx \exp{(-\tau)}$, where τ is the optical thickness of the contaminant layer [Ishimaru 1978]. The intensification $\mathbf{I}_{\boldsymbol{\alpha}}$ is due to the scattering of the contaminants, *i.e.*, the dirt will "gather" light from other directions. So $\mathbf{I}_{\boldsymbol{\alpha}}$ is an integral of the scattering of the outside illumination:

$$\mathbf{I}_{\boldsymbol{\alpha}}(\omega_o) = \int_{\Omega} f_t(\omega_i, \omega_o; \tau, \sigma_a, \sigma_s, g, n) \cdot L_i(\omega_i) d\omega_i, \quad (2)$$

where $L_i(\omega_i)$ is the radiance from the direction ω_i, $f_t(\cdot)$ represents the fraction of light from the direction ω_i scattered to the outgoing direction ω_o, and σ_t, σ_a, g, and n are the absorption coefficient, scattering coefficient, parameter for the phase function, and refractive index, respectively, all of which are material properties of the contaminants. An exact formulation for $f_t(\cdot)$ can be found in [Ishimaru 1978].

We first note that the outside illumination $L_i(\omega_i)$ can be assumed to be the same for all points on camera lenses, since the size of camera lenses is small compared with the scene depths. Moreover, we simplify $f_t(\cdot)$ and assume it is only a function of the optical thickness pattern τ,[3] since (1) most camera lens dirt is optically

---

[2]Similar models have been derived in previous works [Favaro and Soatto 2003; Hasinoff and Kutulakos 2007].

[3]This assumption does not hold if the distribution of $L_i(\omega_i)$ is extremely uneven. We handle this situation in Sec. 4.3.1 and the appendix.
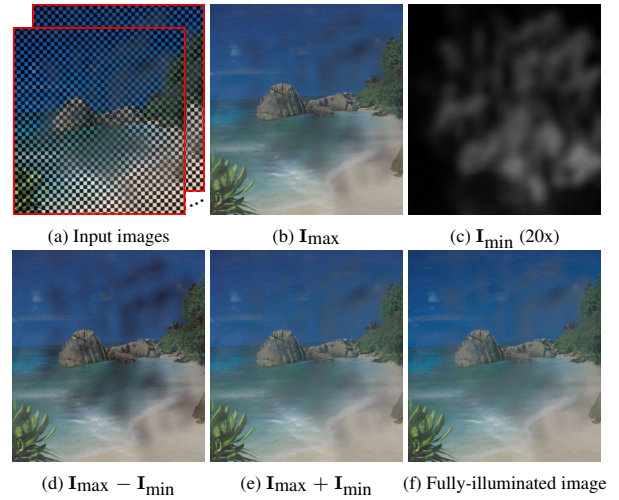

(a) Input images   (b) $\mathbf{I}_{max}$ (c) $\mathbf{I}_{min}$ (20x)

(d) $\mathbf{I}_{max} - \mathbf{I}_{min}$ (e) $\mathbf{I}_{max} + \mathbf{I}_{min}$ (f) Fully-illuminated image

Figure 3: *Validation of Equation (4). (a) A sequence of shifted checkerboard patterns are projected on a scene. (b) The pointwise maximum of the captured images, $\mathbf{I}_{max}$, includes both the attenuation and the scattering. (c) The minimum of the captured images (amplified 20 times for demonstration), $\mathbf{I}_{min}$, directly measures the scattering of the lens dirt. (d) The attenuation can be simply computed as $\mathbf{I}_{max} - \mathbf{I}_{min}$. As shown in (c), the scattering is related only to the attenuation pattern, and not the background scene. (e) shows $\mathbf{I}_{max} + \mathbf{I}_{min}$, and (f) is the image captured when we project a white pattern on the scene. (e) should equal to (f) because the checkerboard patterns turn on half the projector pixels and thus the scattering in (c) is half of the scattering in (f) while the attenuation keeps the same. Indeed, we found (e) and (f) are closely matched with a mean absolute percentage error 0.6%.*

very thin, *i.e.*, τ is close to 0 and thus the variation of $f_t(\cdot)$ caused by ω_i and ω_o is negligible, and (2) compared to the optical thickness pattern τ, the other material properties of lens dirt are less likely be spatially-varying on the camera lens. Specifically, we assume $f_t(\omega_i, \omega_o; \tau, \sigma_a, \sigma_s, g, n) \approx f_t(\tau)$. We have:

$$\mathbf{I}_{\boldsymbol{\alpha}} \approx f_t(\tau) \cdot \int_{\Omega} L_i(\omega_i) d\omega_i = f_t(\tau) \cdot \mathbf{c}, \qquad (3)$$

where $\mathbf{c} = \int_{\Omega} L_i(\omega_i) d\omega_i$ is the aggregate of the outside illumination. Since the attenuation, $\boldsymbol{\alpha}$, is a function of the optical thickness τ, $f_t(\tau)$ can be written instead as a function of $\boldsymbol{\alpha}$:

$$\mathbf{I}_{\boldsymbol{\alpha}}(x, y) = \mathbf{c} \cdot f(\boldsymbol{\alpha}(x, y)), \qquad (4)$$

where we emphasize that both $\mathbf{I}_{\boldsymbol{\alpha}}$ and $\boldsymbol{\alpha}$ are spatially-varying while $\mathbf{c}$ is a fixed vector. The function $f(\boldsymbol{\alpha}(x, y))$ depends on the physical characteristics of the contaminants.

This relationship between $\mathbf{I}_{\boldsymbol{\alpha}}$ and $\boldsymbol{\alpha}$ is important, because it shows that only the aggregate of the outside illumination $\mathbf{c}$ is relevant for intensification. In other words, the intensification $\mathbf{I}_{\boldsymbol{\alpha}}$ is a global effect, and is not directly related to the artifact-free image $\mathbf{I}_0$.

4.1 Model Validation and Simplification

We first validate the simplified model in Equation (4). To do so, we separate the attenuation and scattering components using an approach inspired by [Nayar et al. 2006]. As shown in Fig. 3, we project a sequence of slightly shifted checkerboard patterns and take pictures of the checkerboard modulated scene with a dirty lens camera. Scene points in black squares do not themselves emit light and thus their corresponding pixel intensities are caused by the scattering due to lens dirt. In contrast, pixel intensities of scene points

in white squares include both the light emitted by themselves (after attenuation due to lens dirt), and the light scattered by the lens dirt. The amount of scattered light is fixed since the *integral* of the outside illumination is unchanged. Therefore, letting $\mathbf{I}_{max}$ and $\mathbf{I}_{min}$ denote the pointwise maximum and minimum over all the pictures, we have $\mathbf{I}_{max} = \mathbf{I}_0 \cdot (\boldsymbol{\alpha} * \mathbf{k}) + \mathbf{I}_{\boldsymbol{\alpha}} * \mathbf{k}$ and $\mathbf{I}_{min} = \mathbf{I}_{\boldsymbol{\alpha}} * \mathbf{k}$. Their difference, $\mathbf{I}_{max} - \mathbf{I}_{min}$, is the attenuation. Figure 3 shows the separation results for these components, respectively. As shown, the scattering term, $\mathbf{I}_{min}$, does not relate to the background scene. Moreover, according to our model simplification, $\mathbf{I}_{max} + \mathbf{I}_{min}$ should equal to a fully-illuminated image (*i.e.*, an image captured when we project a white pattern). This is also verified and shown in Figs. 3(d)(f) with a mean absolute percentage error of 0.6%. Therefore, our simplified model is validated: only the aggregate of the outside illumination is relevant to the intensification, *i.e.*, $\mathbf{I}_{\boldsymbol{\alpha}} = \mathbf{c} \cdot f(\boldsymbol{\alpha})$.

Finally, we note that the defocus blur kernel $\mathbf{k}$ can be assumed to be fixed for a given dirty lens camera, since the distance from the contaminant layer to the optical center is usually fixed, and is much smaller than the distance from the scene to the optical center. Thus we can further simplify the model in Equation (1) by defining two variables: the attenuation map $\mathbf{a}(x,y) := \boldsymbol{\alpha}(x,y) * \mathbf{k}(x,y)$ and the intensification map $\mathbf{b}(x,y) := f(\boldsymbol{\alpha}(x,y)) * \mathbf{k}(x,y)$. The model can then be rewritten as:

$$\mathbf{I}(x,y) = \mathbf{I}_0(x,y) \cdot \mathbf{a}(x,y) + \mathbf{c} \cdot \mathbf{b}(x,y), \qquad (5)$$

where $\mathbf{a}(x,y)$ and $\mathbf{b}(x,y)$ are the characteristics for a given camera and dirty pattern and $\mathbf{c}$ is the aggregate of the outside illumination which is scene dependent.

Below, we propose two methods to estimate $\mathbf{a}(x,y)$ and $\mathbf{b}(x,y)$. Note that we only need to estimate these once, after which artifacts can be removed using only a single image.

4.2 Artifact Removal via Calibration

If we are allowed to take several calibration images beforehand, we can use the same idea as above and directly measure its $\mathbf{a}(x,y)$ and $\mathbf{b}(x,y)$ by taking a set of pictures of a structured pattern (*e.g.*, a checkerboard) at different positions:

$$\mathbf{a}(x,y) = \mathbf{I}_{max}(x,y) - \mathbf{I}_{min}(x,y), \quad \mathbf{b}(x,y) = \mathbf{I}_{min}(x,y), \quad (6)$$

where $\mathbf{I}_{max}$ and $\mathbf{I}_{min}$ are the pointwise maximum and minimum over all the pictures, respectively.[4] Ideally, if the black and white pixels in the structured pattern can be swapped exactly in these two pictures, only two pictures are needed. In practice, more pictures can be taken to suppress noise around edges.

In order to remove artifacts from a new image, we only need to estimate the aggregate of the outside illumination $\mathbf{c}$ for the new scene. This is still challenging because $\mathbf{c}$ cannot be simply computed as the average value of the input image. Since the lens itself will usually see a much larger field of view than the image sensor, $\mathbf{c}$ is often greater than the average value of the input image and the information might not be recorded in the captured images.

Instead, we find that $\mathbf{c}$ can be estimated using natural image statistics. Natural images are well known to have strong sparsity in their gradients [Rudin et al. 1992; Levin et al. 2007]. Since dirty-lens artifacts are low frequency (because of the defocus blur), if the artifacts are not completely removed, the recovered image will retain

---

[4]This is true because for the shifted checkerboard images, the target scene radiance is a white image, *i.e.*, $\mathbf{I}_0(x,y) = 1$. The estimate $\mathbf{b}(x,y)$ in Equation (6) has actually been multiplied with the aggregate of the outside illumination of the calibration scene, which is not a problem since the aggregate illumination is a constant.
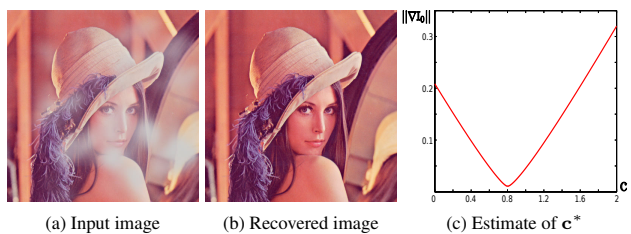

(a) Input image   (b) Recovered image   (c) Estimate of $\mathbf{c}^*$

Figure 4: *Simulation results for solving $\mathbf{c}$ based on Equation (7). (a) A synthetic image with a dirty-lens artifact is generated using the lens dirt pattern extracted from Fig. 3(c). (b) Equation (7) is used to recover an artifact-free image, with a mean absolute percentage error of 0.44% compared to ground truth. (c) A plot of $||\nabla \mathbf{I}_0||_1$ versus $\mathbf{c}$. The optimal estimate, $\mathbf{c}^*$, has a 0.25% error compared to ground truth.*

the low frequency pattern, resulting in a non-sparse image gradient. Therefore, we can estimate $\mathbf{c}$ by enforcing the sparsity of the recovered image's gradient (*i.e.*, minimizing its L_1 norm):

$$\mathbf{c}^* = \arg\min_{\mathbf{c}} ||\nabla \mathbf{I}_0(x,y)||_1, \qquad (7)$$

where

$$\mathbf{I}_0(x,y) = (\mathbf{I}(x,y) - \mathbf{c} \cdot \mathbf{b}(x,y))/\mathbf{a}(x,y). \qquad (8)$$

The optimization is performed for R/G/B channels separately. We first perform a simulation to verify the accuracy of the proposed method. As shown in Fig. 4, we generate a synthetic image with a dirty-lens artifact using the lens dirt pattern extracted in Fig. 3(c). The aggregate of the outside illumination is set to be $\mathbf{c} = 0.80$. Equation (7) is used to estimate $\mathbf{c}^*$ and recover the artifact-free image $\mathbf{I}_0$, as shown in Fig. 4(b). The mean absolute percentage error (MAPE) of the recovered image is 0.44%. The plot in Fig. 4(c) shows $||\nabla \mathbf{I}_0||_1$ at different $\mathbf{c}$. The optimal value of $\mathbf{c}$ is estimated to be 0.802, an error of only 0.25%.

Figure 5 shows experimental results on real images taken with a Canon EOS 20D camera. The camera is equipped with a Canon EF 50mm f/1.8 lens contaminated with house dust powder and fingerprints, thus introducing significant artifacts in the captured images, as shown in Fig. 1(a) and Fig. 5. To estimate the attenuation and scattering terms, we use a structured pattern consisting of black and white vertical stripes printed on a piece of paper. Sixteen pictures are taken at randomly-shifted positions, as shown in Fig. 5(a). The estimated attenuation map $\mathbf{a}(x,y)$ and scattering map $\mathbf{b}(x,y)$ are shown in Fig. 5(b) and Fig. 5(c), respectively. Given this calibration, we use Equation (7) to estimate $\mathbf{c}$ and recover the artifact-free images. Two examples are given in Fig. 5(d)(e)(f). As shown in the insets, the proposed method effectively removes the artifacts caused by the dirty camera lens and reveals more details in the original images. The method works well on photographs taken both indoors and outdoors, and across a large range of outside illuminations.

4.3 Artifact Removal without Calibration

If we do not have access to the camera to perform the above calibration (*e.g.*, for postprocessing existing photographs or videos), we propose a method based on natural image statistics to estimate the attenuation map $\mathbf{a}(x,y)$ and scattering map $\mathbf{b}(x,y)$.

Let us first consider two neighboring pixels p_1 and p_2. Since both $\mathbf{a}(x,y)$ and $\mathbf{b}(x,y)$ are smoothly-varying due to the defocus blur, we have $\mathbf{a}(p_1) \approx \mathbf{a}(p_2)$ and $\mathbf{b}(p_1) \approx \mathbf{b}(p_2)$, and thus we have $\mathbf{I}(p_1) - \mathbf{I}(p_2) \approx (\mathbf{I}_0(p_1) - \mathbf{I}_0(p_2)) \cdot \mathbf{a}(p_1)$. In other words, the magnitude of the image gradient has been attenuated by $\mathbf{a}(p_1)$,
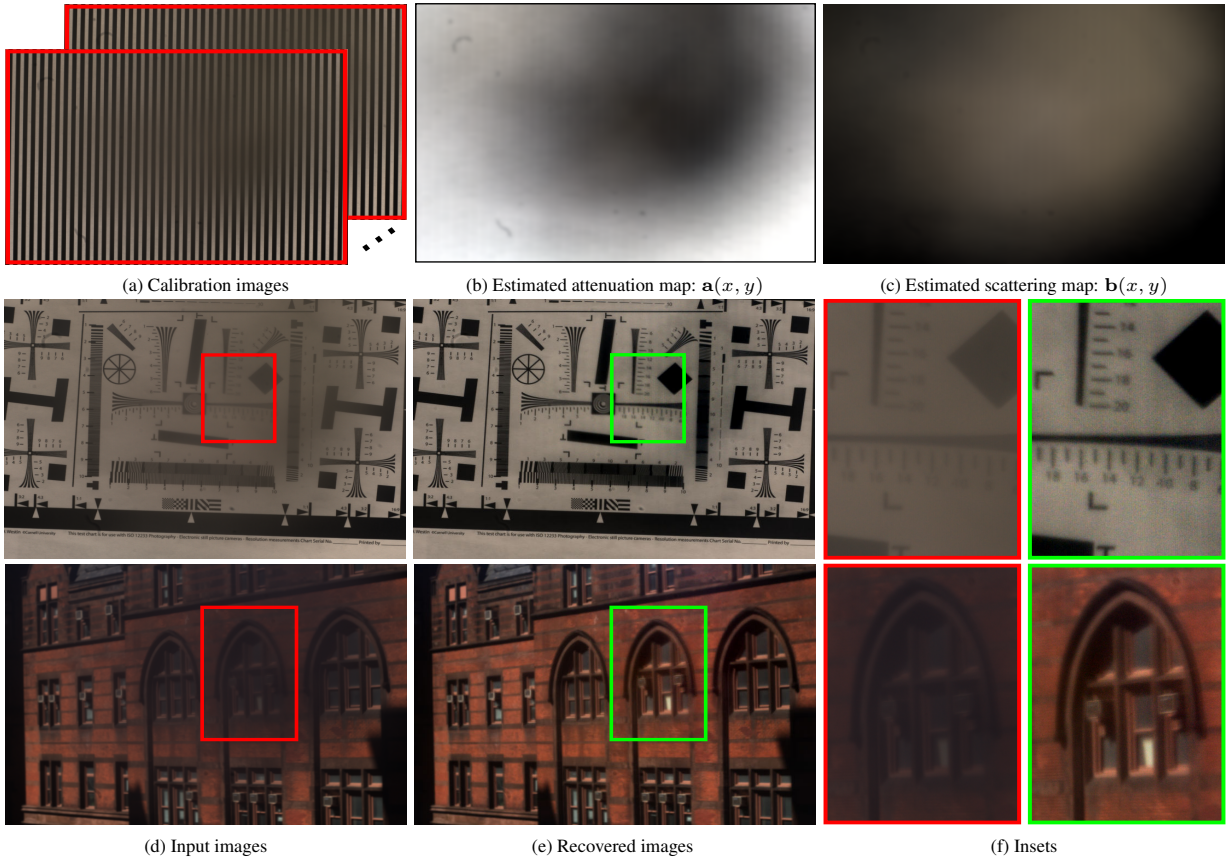
(a) Calibration images   (b) Estimated attenuation map: $\mathbf{a}(x, y)$ (c) Estimated scattering map: $\mathbf{b}(x, y)$

(d) Input images   (e) Recovered images   (f) Insets

Figure 5: *Removal of dirty-lens artifacts via calibration. (a) By taking several pictures of a structured pattern, we can estimate (b) the attenuation map $\mathbf{a}(x, y)$ and (c) the scattering map $\mathbf{b}(x, y)$ for a dirty lens camera. (d) and (e) show that these estimates can be used to remove dirty-lens artifacts for new input images. The optimal estimates of the aggregate outside illumination are $\mathbf{c}^* = [1.37, 1.35, 1.41]$ and $\mathbf{c}^* = [5.48, 5.41, 7.66]$ for the top and the bottom input images, respectively. (f) The insets show that the recovered images reveal more details of the target scenes.*
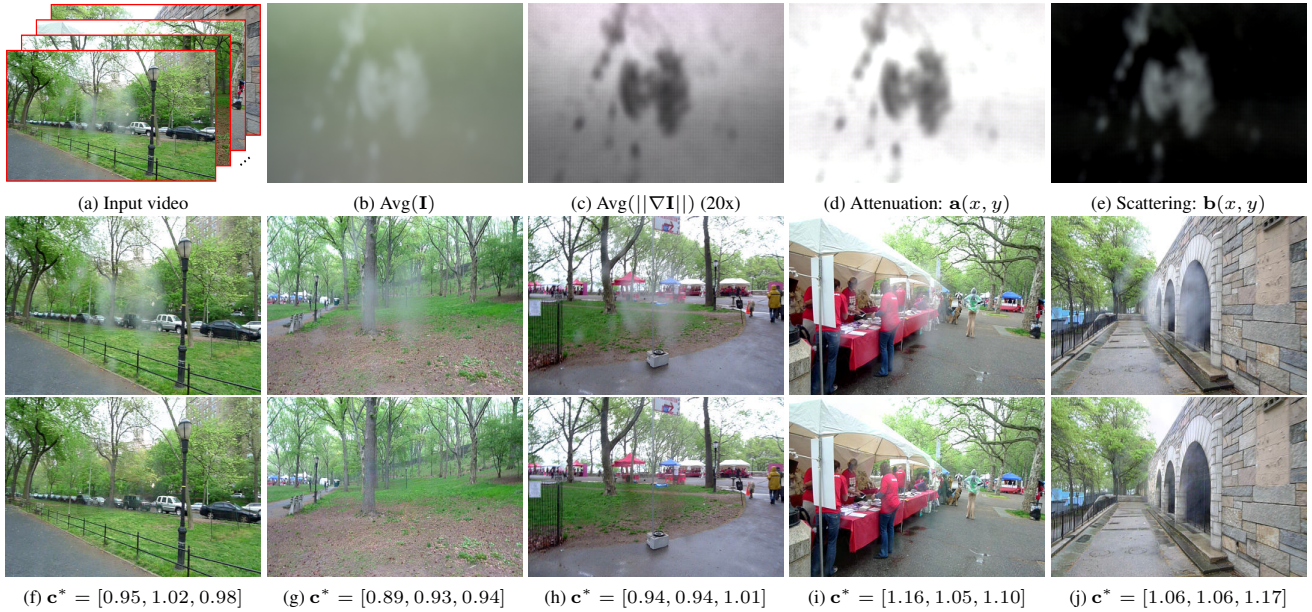

(a) Input video   (b) Avg($\mathbf{I}$)   (c) Avg($||\nabla\mathbf{I}||$) (20x)   (d) Attenuation: $\mathbf{a}(x, y)$ (e) Scattering: $\mathbf{b}(x, y)$

(f) $\mathbf{c}^* = [0.95, 1.02, 0.98]$ (g) $\mathbf{c}^* = [0.89, 0.93, 0.94]$ (h) $\mathbf{c}^* = [0.94, 0.94, 1.01]$ (i) $\mathbf{c}^* = [1.16, 1.05, 1.10]$ (j) $\mathbf{c}^* = [1.06, 1.06, 1.17]$

Figure 6: *Removal of dirty-lens artifacts without calibration. (a) The input is a 5-min long video clip, consisting of 7200 frames. (b) The averaged image over all the frames. (c) The averaged image gradient (amplified 20 times for demonstration) over all the frames. (d) The attenuation map $\mathbf{a}(x, y)$ and (e) the scattering map $\mathbf{b}(x, y)$ can be computed based on Equation (7). (f-j) show examples of artifact removal, where the top row shows the original frames and the bottom row shows the recovered images, along with the estimated $\mathbf{c}^*$.*

which can be stated more formally as:[5]

$$\begin{aligned}
\nabla \mathbf{I} &= \nabla \left(\mathbf{I}_0 \cdot \mathbf{a} + \mathbf{c} \cdot \mathbf{b} \right) \\
&= \left(\nabla \mathbf{I}_0 \right) \cdot \mathbf{a} + \mathbf{I}_0 \cdot \left(\nabla \mathbf{a} \right) + \mathbf{c} \cdot \left(\nabla \mathbf{b} \right) \\
&\approx \left(\nabla \mathbf{I}_0 \right) \cdot \mathbf{a},
\end{aligned}$$

since both $\nabla \mathbf{a} \approx 0$ and $\nabla \mathbf{b} \approx 0$ due to the defocus blur of lens dirt. This relationship holds for every picture taken with the same dirty lens camera, and thus by computing the averaged magnitude of the image gradient over all frames of a video (or a collection of photographs), we have

$$\text{Avg}(|\nabla \mathbf{I}|) \approx \text{Avg}(|\nabla \mathbf{I}_0|) \cdot \mathbf{a}. \qquad (9)$$

where $\text{Avg}(\cdot)$ represents the averaging operation. Similarly,

$$\text{Avg}(\mathbf{I}) \approx \text{Avg}(\mathbf{I}_0) \cdot \mathbf{a} + \bar{\mathbf{c}} \cdot \mathbf{b}, \qquad (10)$$

where $\bar{\mathbf{c}}$ is the averaged aggregate of the outside illumination over all frames and can be absorbed in the estimate of $\mathbf{b}$.

We now rely on natural image statistics to estimate both $\text{Avg}(\mathbf{I}_0)$ and $\text{Avg}(|\nabla \mathbf{I}_0|)$. Since the probability distributions of image intensity and gradient for all pixels are similar to each other, the averaged values (*i.e.*, the expectations) for each of these pixels over a sequence of images should also be similar to each other. This has been shown extensively in previous works [Burton and Moorhead 1987; Torralba et al. 2008; Kuthirummal et al. 2008], where the averaged images are smoothly-varying. This heuristic enables us to estimate $\text{Avg}(\mathbf{I}_0)$ and $\text{Avg}(|\nabla \mathbf{I}_0|)$ from $\text{Avg}(\mathbf{I})$ and $\text{Avg}(|\nabla \mathbf{I}|)$ via a simple iterative polynomial fitting, in RANSAC fashion. More specifically, we model $\text{Avg}(\mathbf{I}_0)$ as a bivariate polynomial, $\sum_{i=0}^{3} \sum_{j=0}^{3} a_{i,j} x^i y^j$, where x, y are *normalized* pixel coordinates in $[-1, 1] \times [-1, 1]$. In the first iteration, we use all the pixels whose values are among the top 50% of $\text{Avg}(\mathbf{I})$ for least-square fitting. We then take the difference between the fitted $\text{Avg}(\mathbf{I}_0)$ and $\text{Avg}(\mathbf{I})$. Those pixels whose residuals are within a threshold (10% of its pixel value) are considered as "inliers" and used to perform least-square fitting in the next iteration. We found that this method converges in $100 \sim 200$ iterations and automatically finds the pixels in the dirty regions as "outliers". The same method is used for estimating $\text{Avg}(|\nabla \mathbf{I}_0|)$.

Therefore, the attenuation map $\mathbf{a}(x, y)$ and scattering map $\mathbf{b}(x, y)$ can be computed from Equations (9) and (10):

$$\mathbf{a} = \text{Avg}(|\nabla \mathbf{I}|)/\text{Avg}(|\nabla \mathbf{I}_0|), \quad \mathbf{b} = \text{Avg}(\mathbf{I}) - \text{Avg}(\mathbf{I}_0) \cdot \mathbf{a}. \ (11)$$

To remove artifacts for individual frames, we use the optimization shown in Equation (7) to estimate the aggregate of the outside illumination, $\mathbf{c}^*$, independently for each frame.

Figure 6 shows experimental results. A Panasonic Lumix DMC3 camcorder was used to take videos at 24fps inside a park on an overcast rainy day. The camcorder's lens is contaminated with fingerprints, dust, and rain drop deposit. As described in Fig. 6(a), a 5-minute clip is used for the computation, resulting in $5 \times 60 \times 24 = 7200$ frames. The image resolution is 1280×720. These frames are used to compute the averaged image and gradient magnitudes, as shown in Figs. 6(b)(c), and to estimate the attenuation map $\mathbf{a}(x, y)$ and scattering map $\mathbf{b}(x, y)$, as shown in Figs. 6(d)(e). Figures 6(f-j) show several examples of artifact removal from images, where the top row shows the original frames and the bottom row shows the recovered images, along with estimates for the aggregate of the outside illumination $\mathbf{c}^*$. The experiments are performed on a 2.4GHZ machine and it takes about 6 seconds for each frame. The input and recovered video clips are submitted as supplementary material, showing that this simple post-processing effectively removes the dirty-lens artifacts.

---

[5]We drop the function parameters (x, y) in order to shorten equations.

(a) Input frames (b) Original algorithm (c) Extended algorithm
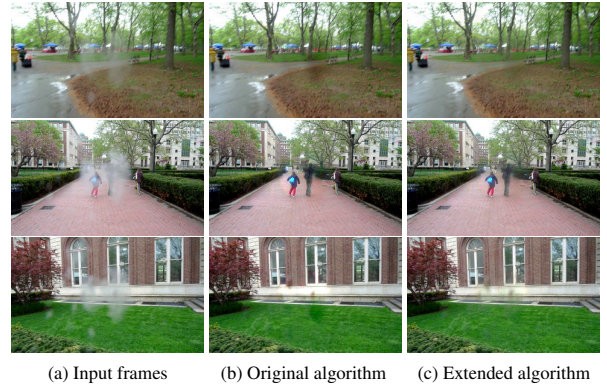
Figure 7: *Post-processing for unevenly distributed lighting. (a) The outside illumination for certain frames is unevenly distributed, causing the intensification $\mathbf{I}_\alpha$ to no longer be independent of the scene texture. (b) Consequently, for these frames, the original algorithm sometimes over-compensates for the dirty regions and causes over-darkening or over-brightening. (c) By modeling the scene texture term in $\mathbf{I}_\alpha$ as shown in Equation (12), we effectively overcome the problem of over-darkening and over-brightening ($w = 0.3$).*

4.3.1 Post-processing for Unevenly Distributed Lighting

For certain frames, the above algorithm might introduce over-darkening and over-brightening artifacts in the dirty regions, as shown in Fig. 7(b). This is caused by the extremely uneven distribution of outside lighting: some dirty regions cover the bright sky, while others cover the ground or bushes that are in shadow. Since many types of contaminants on lenses are forward scattering (*e.g.*, lipids, dermis, dust, droplets) [Ishimaru 1978; Jacques et al. 1987], the unevenly distributed outside lighting causes the intensification $\mathbf{I}_\alpha$ no longer independent of the scene texture (Equation (4)). While the algorithm tries to remove artifacts in the whole image, it sometimes *over-compensates* for dirty regions and causes the over-darkening and over-brightening.

To solve this problem, we extend the above algorithm and explicitly model the scene texture in the intensification (Equation (4)) as

$$\mathbf{I}_\alpha(x, y) = \left(\mathbf{c} + w \cdot \mathbf{I}_0(x, y) \right) \cdot f \left(\boldsymbol{\alpha}(x, y) \right), \qquad (12)$$

where w is a weight coefficient to be estimated. The derivation is shown in the appendix. Accordingly, Equation (8) is modified as

$$\mathbf{I}_0(x, y) = \left(\mathbf{I}(x, y) - \mathbf{c} \cdot \mathbf{b}(x, y) \right)/\left(\mathbf{a}(x, y) + w \cdot \mathbf{b}(x, y) \right). \ (13)$$

The value of w is fixed by trial and error. We found $w = 0.2 \sim 0.8$ gives good results.[6] As shown in Fig. 7(c), this extension effectively overcomes the problem of over-darkening and over-brightening. Compared with the results of the original algorithm shown in Fig. 7(b), although there are slightly more dirty-lens artifacts left in the recovered images, the overall quality is higher.

5 Artifacts Caused by Thin Occluders

As a complement to the dirty-lens artifacts, we now discuss image artifacts caused by a layer of thin occluders between the target scene and the camera. Examples include photographing scenes that can only be observed through meshes such as zoos or sports games, covert surveillance inside a room through a window shutter or tree branches, and underwater imaging within a wired box.

---

[6]Although w can be estimated for each frame together with $\mathbf{c}$ via Equation (7), estimating from individual frames might cause spurious flickering and is time-consuming. Also, w usually varies quite smoothly within the video sequence. Thus we assume w the same value for an input video sequence and search its value by trial and error.
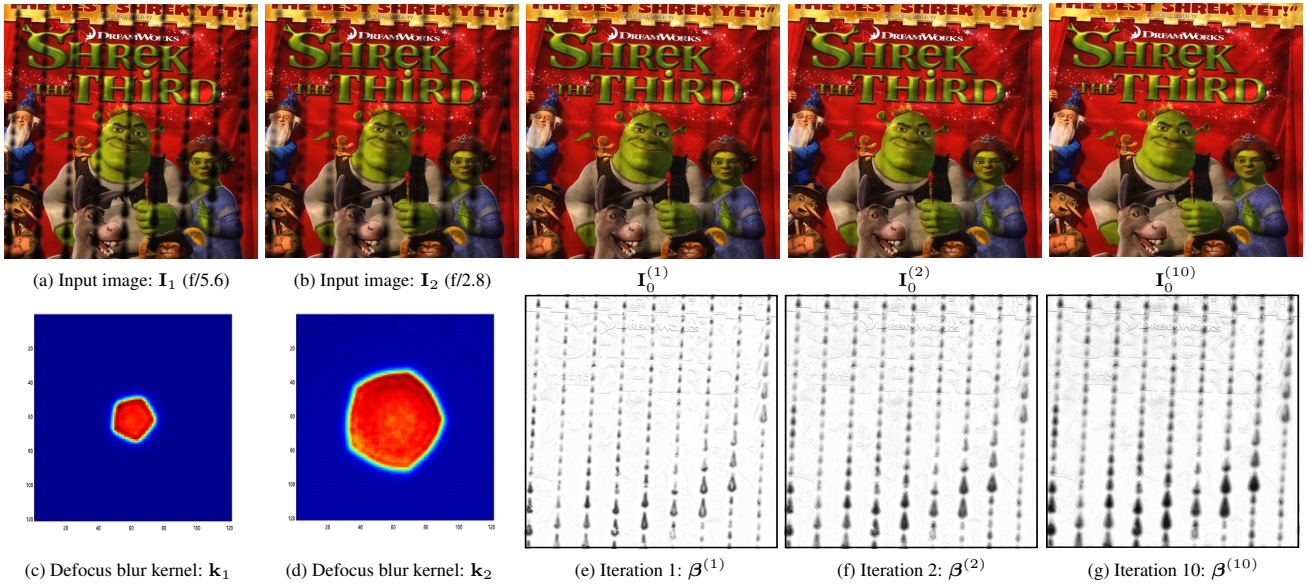
(a) Input image: $\mathbf{I}_1$ (f/5.6)   (b) Input image: $\mathbf{I}_2$ (f/2.8) $\mathbf{I}_0^{(1)}$ $\mathbf{I}_0^{(2)}$ $\mathbf{I}_0^{(10)}$

(c) Defocus blur kernel: $\mathbf{k}_1$ (d) Defocus blur kernel: $\mathbf{k}_2$ (e) Iteration 1: $\boldsymbol{\beta}^{(1)}$ (f) Iteration 2: $\boldsymbol{\beta}^{(2)}$ (g) Iteration 10: $\boldsymbol{\beta}^{(10)}$

Figure 8: *Removal of thin-occluder artifacts from two images. (a) and (b) are the two input images taken with different apertures. (c) and (d) are the two corresponding defocus blur kernels measured beforehand for the given camera at the given depths. (e-g) show the reconstruction results at several iterations, where the top row shows the recovered image and the bottom row shows the estimated occluder pattern.*

Compared with the dirty-lens artifact, although both share the the same image formation model (Equation (1)), there are several differences in this case: (1) we can no longer treat $\boldsymbol{\alpha} * \mathbf{k}$ and $\mathbf{I}_\alpha * \mathbf{k}$ as single variables, since the defocus blur kernel $\mathbf{k}$ might change according to the depths of the occluder layer and the scene; (2) the intensification $\mathbf{I}_\alpha$ is caused not by the scattering of contaminants but by the reflection of the occluders, and thus we do not have a direct relationship between $\mathbf{I}_\alpha$ and $\boldsymbol{\alpha}$.

In this paper we consider a special case where the radiance from the occlusion layer itself is negligible compared to radiance from the background (*i.e.*, $\mathbf{I}_\alpha \approx 0$). Example scenarios include, for example, pictures taken from a room where the outdoor light is much stronger than the light inside the room, or for security videos taken by cameras through dirty protectors, where the dirt/mud is so dense that it completely obstructs the incoming light and will not scatter light from other directions.

5.1 Known Depths: Artifact Removal with Two Images

In the simplest case, if the depths of the occluding layer and the scene are known, we show that image artifacts can be removed using two images with different apertures. In the two images, the background scene is always in focus while the occluder layer will have different amounts of defocus blur. More specifically, the image formation model in this case can be simplified to

$$\mathbf{I}_1 = \mathbf{I}_0 \cdot (\boldsymbol{\alpha} * \mathbf{k}_1), \quad \mathbf{I}_2 = \mathbf{I}_0 \cdot (\boldsymbol{\alpha} * \mathbf{k}_2),$$

where the defocus blur kernels $\mathbf{k}_1$ and $\mathbf{k}_2$ are known, given the depths of the scene and the occluder layer.

We propose a simple yet effective method to solve the equations based on fixed-point iteration methods [Saad 1996; Zeid 1985]. Let us first further simplify the above equations by introducing a new variable $\boldsymbol{\beta} = \boldsymbol{\alpha} * \mathbf{k}_1$:

$$\mathbf{I}_1 = \mathbf{I}_0 \cdot \boldsymbol{\beta}, \quad \mathbf{I}_2 = \mathbf{I}_0 \cdot (\boldsymbol{\beta} * \mathbf{k}),$$

where $\mathbf{k}$ is the blur kernel such that $\boldsymbol{\beta} * \mathbf{k} = \boldsymbol{\alpha} * \mathbf{k}_2$. There is no need to compute $\mathbf{k}$ explicitly, since $\boldsymbol{\beta} * \mathbf{k}$ can be evaluated using the Fourier transform as follows:

$$\boldsymbol{\beta} * \mathbf{k} = \boldsymbol{\alpha} * \mathbf{k}_2 = \mathrm{F}^{-1}\left(\frac{\mathrm{F}^*(\mathbf{k}_1)\mathrm{F}(\mathbf{k}_2)}{\mathrm{F}^*(\mathbf{k}_1)\mathrm{F}(\mathbf{k}_1) + \lambda} \cdot \mathrm{F}(\boldsymbol{\beta})\right) \quad (14)$$

where $\mathrm{F}(\cdot)$, $\mathrm{F}^{-1}(\cdot)$, and F^* are the forward, inverse, and conjugate Fourier transforms, respectively, and λ is a constant commonly defined in Wiener filters representing the noise to signal ratio.[7] Our proposed algorithm solves the above equations using the following fixed-point iterations with the Jacobi preconditioner [Saad 1996]:

$$\tilde{\boldsymbol{\beta}}^{(i+1)} = (1 - w) \cdot \boldsymbol{\beta}^{(i)} + w \cdot \frac{\mathbf{I}_1}{\mathbf{I}_2} \cdot (\boldsymbol{\beta}^{(i)} * \mathbf{k}), \quad (15)$$

where the Jacobi preconditioner matrix $w = (1 - k_c \mathbf{I}_1/\mathbf{I}_2)^{-1}$ and k_c is the value of the center point in the blur kernel $\mathbf{k}$. For each iteration, given the estimate of $\boldsymbol{\beta}^{(i)}$, the artifact-free image can be easily computed as $\mathbf{I}_0^{(i)} = \mathbf{I}_1/\boldsymbol{\beta}^{(i)}$.

The convergence of the above iteration depends on the defocus blur kernels $\mathbf{k}_1$ and $\mathbf{k}_2$, as well as the content of the artifact-free image, $\mathbf{I}_0$. A more detailed discussion can be found in Sec. 6. In practice, we use a simple strategy to ensure the convergence of the algorithm by updating only pixels where the reconstruction error is decreasing at each iteration. More specifically, $\boldsymbol{\beta}$ is updated as follows:

$$\boldsymbol{\beta}^{(i+1)} = \begin{cases} \tilde{\boldsymbol{\beta}}^{(i+1)}, & \text{if } \mathrm{Err}(\tilde{\boldsymbol{\beta}}^{(i+1)}) < \mathrm{Err}(\boldsymbol{\beta}^{(i)}) \\ \boldsymbol{\beta}^{(i)}, & \text{otherwise,} \end{cases} \quad (16)$$

where $\mathrm{Err}(\boldsymbol{\beta})$ denotes the reconstruction error, and is defined as $\mathrm{Err}(\boldsymbol{\beta}) = ||\mathbf{I}_1 - \mathbf{I}_0 \cdot \boldsymbol{\beta}||^2 + ||\mathbf{I}_2 - \mathbf{I}_0 \cdot (\boldsymbol{\beta} * \mathbf{k})||^2$.

Figure 8 shows an experimental result for this simple scenario. A Canon EOS 20D camera was used to take pictures of the movie poster through a door curtain with a Canon EF 50mm f/1.8 lens. The distance between the camera and the door is 150mm and the distance between the door and the poster is 210mm. We take two images at f/2.8 and f/5.6, respectively, as shown in Fig. 8(a)(b). Figures 8(c)(d) show the corresponding defocus blur kernels $\mathbf{k}_1$ and $\mathbf{k}_2$ measured with a planar array of point light sources [Zhou and Nayar 2009]. Although a pillbox function could be used instead to avoid this calibration, we found the measured kernels usually give more accurate results. Figures 8(e)(f)(g) show the reconstruction results for several iterations, including both the recovered image $\mathbf{I}_0^{(i)}$ and
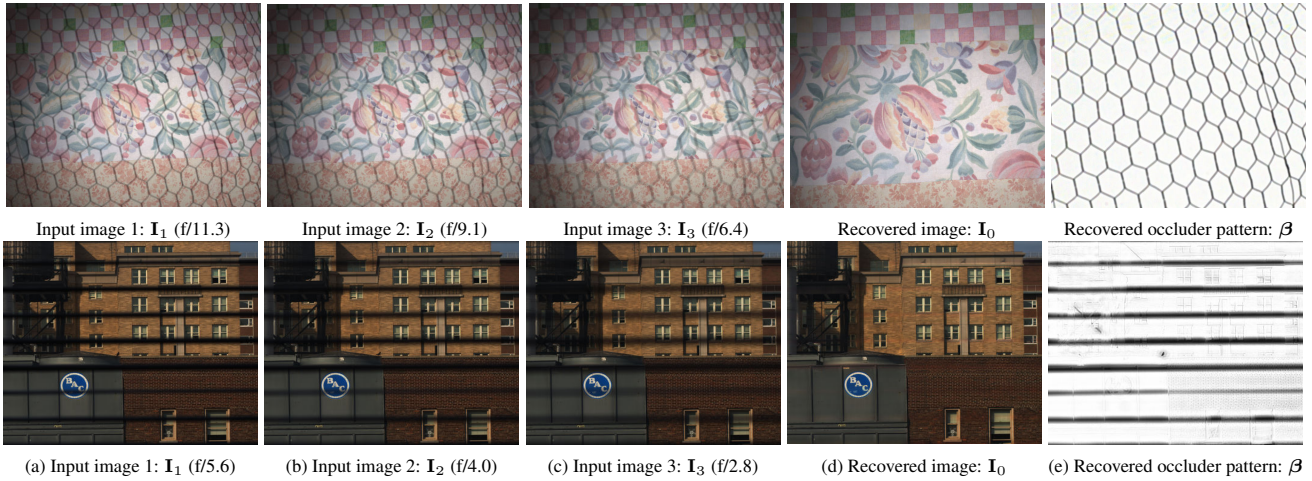
---

[7] In all our experiments we set $\lambda = 0.01$.

| Input image 1: $\mathbf{I}_1$ (f/11.3) | Input image 2: $\mathbf{I}_2$ (f/9.1) | Input image 3: $\mathbf{I}_3$ (f/6.4) | Recovered image: $\mathbf{I}_0$ | Recovered occluder pattern: $\boldsymbol{\beta}$ |

| (a) Input image 1: $\mathbf{I}_1$ (f/5.6) | (b) Input image 2: $\mathbf{I}_2$ (f/4.0) | (c) Input image 3: $\mathbf{I}_3$ (f/2.8) | (d) Recovered image: $\mathbf{I}_0$ | (e) Recovered occluder pattern: $\boldsymbol{\beta}$ |

Figure 9: *Removal of thin-occluder artifacts from three images. (a)(b)(c) are the three input images taken with different apertures. The proposed algorithm is used to remove the artifacts from the input images without knowing the depth of the occluder layer. (d) are the recovered images. (e) are the estimated occluder patterns.*

the estimated occluder pattern $\boldsymbol{\beta}^{(i)}$. As can be seen in Fig. 8(g), the proposed algorithm effectively removes the thin-occluder artifacts from the input images. In our experiments, we found the algorithm converges within $10 \sim 20$ iterations for most scenes. For processing a 1800×1500 color image, it will take about 3 minutes on a 2.4GHz machine.

5.2 Unknown Depths: Removal with Three Images

If the depth of the scene or the occluding layer is unknown, the problem can be solved by using one more image with a different aperture size. This third aperture image is used to find the optimal scale of the defocus blur for each of the blocks of the input images. More specifically, let $\mathbf{I}_1, \mathbf{I}_2, \mathbf{I}_3$ be the three images taken with different apertures, in which the background scene is always in focus. Their defocus blur kernels for the occluder layer are $\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3$, respectively. Since the depths are unknown, $\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3$ are similar up to a scale transformation. Let $\mathbf{k}(s)$ represent the scaled defocus blur kernel $\mathbf{k}$ with scale s. For any given scale s, we take two of the three images, scale the defocus blur kernels accordingly, and solve for $\boldsymbol{\beta}$ and $\mathbf{I}_0$ based on Equation (15). Let $\hat{\boldsymbol{\beta}}(\mathbf{I}_1, \mathbf{I}_2, s)$ and $\hat{\mathbf{I}}_0(\mathbf{I}_1, \mathbf{I}_2, s)$ denote the estimated $\boldsymbol{\beta}$ and $\mathbf{I}_0$ using the first and the second images under scale s. Using the estimated $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{I}}_0$ as well as the scaled defocus blur kernel $\mathbf{k}_3(s)$, we can then reconstruct the third image and compute the error. The optimal scale s^* can be found with a simple 1D optimization as the value that yields the minimum error:

$$s^* = \arg\min_s ||\mathbf{I}_3 - \hat{\mathbf{I}}_0(\mathbf{I}_1, \mathbf{I}_2, s) \cdot < \hat{\boldsymbol{\beta}}(\mathbf{I}_1, \mathbf{I}_2, s), \mathbf{k}_3(s) >||.$$

where the operation $< \cdot >$ is defined in Equation (14) with $\mathbf{k}_2$ substituted with $\mathbf{k}_3(s)$. This 1D optimization is performed block-wise (with overlapped-regions) for the entire image.

Figure 9 shows two experimental results for this case. The first example, as shown in the top row of the figure, is an indoor scene where we take pictures of a board through a hexagon wired mesh. The hexagons on the right boundary are of smaller size, indicating they are further away from the camera. In the second example, as shown in the bottom row of the figure, we take pictures of outdoor scenes inside a room through the window shutter in order to mimic the situation of covert surveillance. Figures 9(a)(b)(c) show the three images taken with different apertures: f/11.3, f/9.1, and f/6.4 for the first example, and f/5.6, f/4.0, and f/2.8 for the second example, respectively. The image resolution is 1800×1500

pixels. The images are uniformly divided into 12×12 blocks and each block is of size 450×375 pixels (*i.e.*, two neighboring blocks overlap in 1/3 of their area). Given the three input images, we used our proposed algorithm to find the optimal scales for blocks of the images and remove the thin-occluder artifacts. The total run time is $10 \sim 15$ mintues on a 2.4GHz machine. Figures 9(d)(e) show the corresponding recovered artifact-free images and the estimated occluder patterns. These results show that the proposed algorithm can effectively remove these artifacts from images without knowing the depths of the occluders or the scene.

6 Limitations and Discussion

Noise amplification: In our approach for the removal of dirty-lens artifacts, the artifact-free image $\mathbf{I}_0$ is computed as $(\mathbf{I} - \mathbf{c} \cdot \mathbf{b})/\mathbf{a}$. Since $\mathbf{a}$ is less than 1 where there is lens dirt, the amount of noise in those recovered pixels will be amplified by $1/\mathbf{a}$ times. For example, some color noise can be observed in the regions with the greatest amount of lens dirt in Fig. 1(b), Fig. 5, and Fig. 6. In fact, this is a common and fundamental problem for almost all pointwise image enhancement operations [Treibitz and Schechner 2009]. To suppress this amplification of noise during the process of removing dirty-lens artifacts, standard high-frequency preserving operations such as bilateral filtering can be performed afterwards. Moreover, it is recommended to take high dynamic range images when possible.

Choosing the apertures for thin-occluder artifact removal: Equation (15) relies on the difference between the defocus blur of the occluder layer in the two input images. The effectiveness of the algorithm depends on both the defocus blur kernels and the occluder pattern. A theoretical analysis is difficult to derive. Instead, we performed a set of numerical simulations and found several heuristic rules for choosing the appropriate apertures.

Let us assume the occluder pattern consists of repeated stripes, in which the width of each stripe is denoted by d and the gap between two neighboring stripes is d_g. We assume the defocus blur kernels $\mathbf{k}_1$ and $\mathbf{k}_2$ are pillbox functions.[8] Let r_1 and r_2 denote the radius of the two pillbox functions, where $r_1 < r_2$. Three heuristic rules we found effective in practice are:

- $r_2 \geq d/2$: to ensure every scene point is partially visible in at least one of the input images. Otherwise, there will be "black

---

[8]This is a better approximation of the real PSF than the more commonly used gaussian functions.

(a) Input image with lens flare   (c) Input image through thick occluders

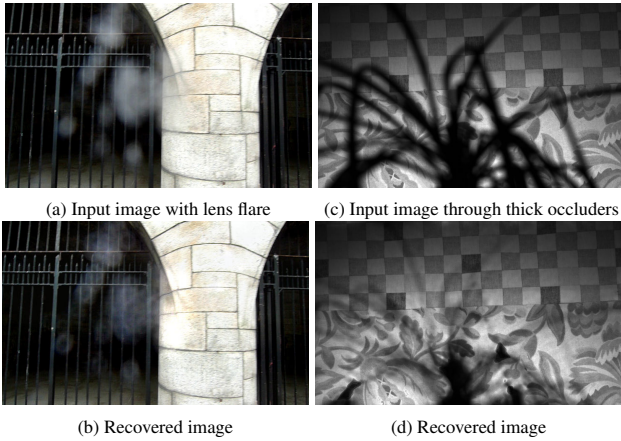(b) Recovered image   (d) Recovered image

Figure 10: *Two failure cases. Case 1: (a) lens flares are caused not only by scattering but also inter-reflection and diffraction within camera lens systems, and thus have strong directionality in the scattering map. (b) Our method can only partially remove the artifacts from the input image. Case 2: (c) for thick occluders, part of the background scene will be blocked in all the input images, and thus cannot be recovered in a pointwise manner (d).*

holes" in the recovered image, as shown in Fig. 10(d).

- $r_2 \le (d + d_g)/2$: to ensure that the highest frequency in the occluder pattern will be within the first band of the spectrum of the pillbox function $\mathbf{k}_2$. For situations that violate this rule (*e.g.*, looking through dense curtains or cloths), low-frequency (*i.e.*, smoothly-varying) artifacts will remain in both the recovered image and the estimated occluder pattern.

- $r_2 \ge \sqrt{2}r_1$: to ensure the two defocus kernels are sufficiently different from each other to prevent ringing artifacts. This corresponds to setting the two apertures at least 1 stop apart.

These limitations are essentially due to the poor behavior of the pillbox function in the frequency domain, a common problem with conventional digital cameras. Recent advances in computational photography [Levin et al. 2007; Zhou and Nayar 2009] show that by using coded apertures one can expect to overcome this problem.

Failure cases: lens flare and thick occluders Figure 10 shows two typical failure cases. For the removal of dirty-lens artifacts, if camera lenses are exposed to strong light sources such as the sun, lens flares will appear in the image. Since lens flares usually have strong directionality, our assumption that the scattering $\mathbf{b}$ is independent of the background is violated, and thus as shown in Fig. 10(a) and (b), our proposed method cannot completely remove the artifacts. For thin-occluder artifacts, if the chosen apertures are too small compared with the size of the occluders, there will be black holes left in the recovered image, as shown in Fig. 10(c)(d), since these regions are completely obstructed in all input images. Image inpainting methods might be used to fill these holes.

7   Conclusions

In this work, we have studied image artifacts caused by taking pictures with dirty camera lenses and by imaging through a layer of thin occluders. Both artifacts can be modeled as an intermediate layer between the camera lens and the target scene that both *attenuates* the background scene and *intensifies* the resulting image (via either scattering or surface reflection). Based on this image formation model, we devised several methods to remove the artifacts from images, either with a simple calibration step or taking multiple images with different apertures if we have access to the imaging systems, or by relying on the statistics of natural images

for post-processing existing images or videos. Experimental results on various scenes show the effectiveness of our methods.

We believe that given the simplicity of the proposed algorithms (especially the removal of lens dust), they will have wide applications in image and video post-processing. For the removal of thin occluders, its current limitation of requiring multiple images could be alleviated by using coded apertures and multiple co-located sensors or mirrors in the future.

Appendix

Here we show the derivation of Equation (12) for unevenly distributed lighting. When the distribution of $L_i(\omega_i)$ is extremely uneven, $f_t(\cdot)$ in Equation (2) can no longer be assumed to be independent of the incident and outgoing directions, ω_i and ω_o, even if the optical thickness τ is small. To take into account this angular dependency of $f_t(\cdot)$, we factorize the function $f_t(\omega_i, \omega_o; x)$ for an optically-thin layer [Gu et al. 2007] as $f_t(\omega_i, \omega_o; \tau) \approx f_1(\tau) \cdot f_2(\omega_i, \omega_o)$, where $f_1(\tau)$ is a function of τ only, representing spatially-varying optical thickness, and $f_2(\omega_i, \omega_o)$ is the phase function.

For isotropic scattering, $f_2(\omega_i, \omega_o) = $ constant, and thus $f_t(\cdot)$ reduces to a function of τ only as before. However, many types of contaminants on lenses are mainly forward scattering (*e.g.*, lipids, dermis, dust, droplets) [Ishimaru 1978; Jacques et al. 1987]. We thus assume the phase function of lens dirt as a simplified Delta-Eddington [Joseph et al. 1976] function $f_2(\omega_i, \omega_o) \approx (1 - w) + w \cdot \delta(\omega_i - \omega_o)$, where w is the weight coefficient. Substituting this into Equation (2), we have:

$$\mathbf{I_\alpha}(\omega_o) = f_1(\tau) \cdot (\mathbf{c} + w \cdot \mathbf{I}_0(\omega_o)),$$

where each outgoing direction ω_o corresponds to one pixel (x, y). Note that since the Delta-Eddington function is an approximation of the real scattering (which often results in a lower-frequency of the incident light $L_i(\omega_i)$), the weight w will not only be determined by the material properties of lens dirt, but also it will change according to the distribution of outside illumination $L_i(\omega_i)$.

References

BERTALMIO, M., SAPIRO, G., CASELLES, V., AND BALLESTER, C. 2000. Image inpainting. In *Proceedings of SIGGRAPH 2000*, 417–424.

BURTON, G. J., AND MOORHEAD, I. R. 1987. Color and spatial structured in natural scenes. *Applied Optics 26*, 1, 157–170.

DURAND, F., MCGUIRE, M., MATUSIK, W., PFISTER, H., AND HUGHES, J. F. 2005. Defocus video matting. *ACM Transactions on Graphics (SIGGRAPH) 24*, 3, 567–576.

EFROS, A. A., AND FREEMAN, W. T. 2001. Image quilting for texture synthesis and transfer. In *Proceedings of SIGGRAPH 2001*, 341–346.

FATTAL, R. 2008. Single image dehazing. *ACM Transactions on Graphics (SIGGRAPH) 27*, 3, 72:1–72:9.

FAVARO, P., AND SOATTO, S. 2003. Seeing beyond occlusions (and other marvels of a finite lens aperture). In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 579–586.

GU, J., RAMAMOORTHI, R., BELHUMEUR, P., AND NAYAR, S. 2007. Dirty glass: Rendering contamination on transparent surfaces. In *Eurographics Symposium on Rendering (EGSR)*, 159–170.

HANSEN, P. C., NAGY, J. G., AND P.O'LEARY, D. 2006. *Deblurring Images: Matrices, Spectra, and Filtering*. Society of Industrial and Applied Mathematics (SIAM).

HASINOFF, S. W., AND KUTULAKOS, K. N. 2007. A layer-based restoration framework for variable-aperture photography. In *IEEE International Conference on Computer Vision (ICCV)*.

HE, K., SUN, J., AND TANG, X. 2009. Single image haze removal using dark channel prior. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

ISHIMARU, A. 1978. *Wave Propagation and Scattering in Random Media*. Academic Press, New York.

JACQUES, S. L., ALTER, C. A., AND PRAHL, S. A. 1987. Angular dependence of HeNe laser light scattering by human dermis. *Lasers Life Science 1*, 309–333.

JOSEPH, J. H., WISCOMBE, W. J., AND WEINMAN, J. A. 1976. The Delta-Eddington approximation for radiative flux transfer. *Journal of the Atmospheric Sciences 33*, 12, 2452–2459.

KOMODAKIS, N., AND TZIRITAS, G. 2006. Image completion using global optimization. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 442–452.

KOREBAN, F., AND SCHECHNER, Y. Y. 2009. Geometry by deflaring. In *IEEE International Conference on Computational Photography (ICCP)*.

KUTHIRUMMAL, S., AGARWALA, A., GOLDMAN, D. B., AND NAYAR, S. K. 2008. Priors for large photo collections and what they reveal about cameras. In *Proceedings of European Conference on Computer Vision (ECCV)*, 74–87.

LEVIN, A., ZOMET, A., AND WEISS, Y. 2003. Learning how to inpaint from global image statistics. In *IEEE International Conference on Computer Vision (ICCV)*, 305–312.

LEVIN, A., FERGUS, R., DURAND, F., AND FREEMAN, W. T. 2007. Image and depth from a conventional camera with a coded aperture. *ACM Transactions on Graphics (SIGGRAPH) 26*, 3, 70:1–70:10.

LIU, Y., BELKINA, T., HAYS, J. H., AND LUBLINERMAN, R. 2008. Image de-fencing. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

MCCLOSKEY, S., LANGER, M., AND SIDDIQI, K. 2007. Automated removal of partial occlusion blur. In *Asian Conference on Computer Vision (ACCV)*, 271–281.

NARASIMHAN, S., AND NAYAR, S. 2003. Contrast restoration of weather degraded images. *IEEE Transactions on Pattern Analysis and Machine Intelligence 25*, 6, 713–724.

NAYAR, S. K., KRISHNAN, G., GROSSBERG, M. D., AND RASKAR, R. 2006. Fast separation of direct and global components of a scene using high frequency illumination. *ACM Transactions on Graphics (SIGGRAPH) 25*, 3, 935–944.

RASKAR, R., AGRAWAL, A., WILSON, C. A., AND VEER-ARAGHAVAN, A. 2008. Glare aware photography: 4D ray sampling for reducing glare effects of camera lenses. *ACM Transactions on Graphics (SIGGRAPH) 27*, 3, 56:1–56:10.

RUDIN, L. I., OSHER, S., AND FATEMI, E. 1992. Nonlinear total variation based noise removal algorithms. *Physica D 60*, 1-4, 259–268.

SAAD, Y. 1996. *Iterative Methods for Sparse Linear Systems*. PWS Publishing Company.

SCHECHNER, Y. Y., AND KARPEL, N. 2005. Recovery of underwater visibility and structure by polarization analysis. *IEEE Journal of Oceanic Engineering 30*, 3, 570–587.

SCHECHNER, Y. Y., NARASIMHAN, S. G., AND NAYAR, S. K. 2001. Instant dehazing of images using polarization. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 325–332.

SCHECHNER, Y., NARASIMHAN, S., AND NAYAR, S. 2003. Polarization-based vision through haze. *Applied Optics, Special issue 42*, 3, 511–525.

SUN, J., YUAN, L., JIA, J., AND SHUM, H.-Y. 2005. Image completion with structure propagation. *ACM Transactions on Graphics (SIGGRAPH) 24*, 3, 861–868.

TALVALA, E.-V., ADAMS, A., HOROWITZ, M., AND LEVOY, M. 2007. Veiling glare in high dynamic range imaging. *ACM Transactions on Graphics (SIGGRAPH) 26*, 3, 37:1–37:10.

TORRALBA, A., FERGUS, R., AND FREEMAN, W. T. 2008. 80 million tiny images: A large data set for nonparametric object and scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence 30*, 11, 1958–1970.

TREIBITZ, T., AND SCHECHNER, Y. Y. 2009. Recovery limits in pointwise degradation. In *IEEE International Conference on Computational Photography (ICCP)*.

VAISH, V., LEVOY, M., SZELISKI, R., ZITNICK, C. L., AND KANG, S. B. 2006. Reconstructing occluded surfaces using synthetic apertures: Stereo, focus and robust measures. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2331–2338.

WATANABE, M., AND NAYAR, S. 1996. Minimal operator set for passive depth from defocus. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 431–438.

WILLSON, R. G., MAIMONE, M. W., JOHNSON, A. E., AND SCHERR, L. M. 2005. An optical model for image artifacts produced by dust particles on lenses. In *International Symposium on Artificial Intelligence, Robtics and Automation in Space (i-SAIRAS)*.

ZEID, I. 1985. Fixed-point iteration to nonlinear finite element analysis. part i: Mathematical theory and background. *International Journal for Numerical Methods in Engineering 21*, 2027–2048.

ZHOU, C., AND LIN, S. 2007. Removal of image arifacts due to sensor dust. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

ZHOU, C., AND NAYAR, S. 2009. What are good apertures for defocus deblurring? In *IEEE International Conference on Computational Photography (ICCP)*.