Veiling Glare in High Dynamic Range Imaging

Eino-Ville Talvala[1] Andrew Adams[2] Mark Horowitz[1] Marc Levoy[2]

[1]Electrical Engineering Department
Stanford University

[2]Computer Science Department
Stanford University

(a) (b)

Figure 1: *Two HDR captures of a strongly backlit scene, tonemapped for printing. The camera is a Canon 20D. (a) Backlighting produces veiling glare in the camera body and lens, visible as a loss of contrast in the foreground objects. (b) By interposing a structured occlusion mask between the camera and the scene, capturing a sequence of HDR images under different translations of the mask, and applying the algorithm described in Section 4.3, we produce an image with substantially reduced glare.*

Abstract

The ability of a camera to record a high dynamic range image, whether by taking one snapshot or a sequence, is limited by the presence of veiling glare - the tendency of bright objects in the scene to reduce the contrast everywhere within the field of view. Veiling glare is a global illumination effect that arises from multiple scattering of light inside the camera's body and lens optics. By measuring separately the direct and indirect components of the intra-camera light transport, one can increase the maximum dynamic range a particular camera is capable of recording. In this paper, we quantify the presence of veiling glare and related optical artifacts for several types of digital cameras, and we describe two methods for removing them: deconvolution by a measured glare spread function, and a novel direct-indirect separation of the lens transport using a structured occlusion mask. In the second method, we selectively block the light that contributes to veiling glare, thereby attaining significantly higher signal-to-noise ratios than with deconvolution. Finally, we demonstrate our separation method for several combinations of cameras and realistic scenes.

CR Categories: I.4.1 [Image Processing and Computer Vision]: Digitization and Image capture—Radiometry

Keywords: Computational photography, HDR imaging, Veiling glare, Global illumination, Structured occlusion mask

1 Introduction

The goal of high dynamic range (HDR) imaging is to accurately capture scenes with a greater luminance range than traditional cameras can capture. This imaging technique is becoming increasingly important in science, medicine, and photography [Reinhard et al. 2006].

However, the capture of a scene's true luminance is limited by the recording equipment used. For a digital camera, a point light source in focus should contribute light to only a single sensor pixel. In reality, it will also contribute some light to all other sensor pixels, with an intensity that depends on the quality of the lens and the distance from the correct pixel (Figure 4). The 2D function describing the intensities is known as the glare spread function (GSF) of the lens.

The center of the GSF is the desired direct illumination path through the lens. The remainder is referred to as veiling glare and lens flare. Veiling glare is caused by scattering in lens elements, reflections off the lens barrel, reflections off lens surfaces, and reflections from the sensor surface itself. Effects that create sharp image artifacts are categorized as lens flare, and include ghost images of the lens aperture and streaks created by diffraction off the aperture leaves. Most of these effects can be seen in Figure 3. If the light source is extended, the lens flare becomes blurred out and becomes veiling glare. Glare is not always caused by visible scene features; it can also be caused by light entering the device from outside of its field of view. This type of glare, however, can usually be avoided by using a lens hood. In general, more inexpensive lenses exhibit more glare.

If a scene is captured with a high-quality lens and camera, with an exposure set so that no part of the sensor is saturated, the level of glare will be below the 8-12 bit dynamic range of the camera, and hence invisible. But multi-exposure HDR capture [Mann and Picard 1995; Debevec and Malik 1997] extends the effective dynamic

range of the camera, and can therefore be limited by veiling glare rather than by the range of exposures available [McCann and Rizzi 2007]. As HDR imaging becomes more popular, it is therefore important to address the problem of veiling glare

We observe that veiling glare is a global illumination effect occurring inside the camera body and lens. Based on this observation, we explore two methods for reducing this unwanted global light transport: deconvolution, and a novel method using a structured occlusion mask. If we assume the glare spread function is shift invariant, then glare formation amounts to a very mild blur of the original image. This leads to a well-conditioned deconvolution problem. We show, however, that with strong veiling glare, deconvolution performs poorly due to limited camera precision. At worst, true scene content can disappear entirely under the quantization floor of the sensor. In general, any glare removal method that merely post-processes a standard HDR capture will fail because of camera precision limits.

Our second method places a flat occluder between the camera and scene, perforated with holes. The regions of the sensor that image through holes record the direct component of light, plus the veiling glare. Adjacent sensor regions that image only the occluder record only glare from the adjacent holes. We form an estimate for the glare in the holes by interpolating the glare values from occluded regions across the holes. Then we subtract the glare estimate from the image, resulting in a glare-free estimate for the hole regions. We translate the occluder slightly and repeat the process, until we have glare-free estimates for the whole scene. This process is inspired by recent work on direct/global separation [Nayar et al. 2006], where light transport is decomposed into a direct and global transport image. For a lens system this corresponds to the desired image and the glare image, respectively. Figure 2 illustrates the parallels between the two problems.

As in the work of Nayar et al.[2006], we subtract two images of comparable magnitude. In situations involving strong glare, the subtraction may produce a noisy estimate of the direct component. We can improve our noise performance by reducing the fraction of the mask that is unoccluded. Since glare is a low spatial frequency phenomenon, it will be attenuated in proportion. With this refinement, we remove the majority of the glare-producing light before it reaches the camera optics, and thereby improve our signal-to-noise ratio.

We begin, in Section 2, with a discussion of related work in optics, medical imaging, astronomy, and computer graphics. In Section 3, we describe our method for recording point-source glare and flare, and we measure glare spread functions for several cameras. We then discuss removing glare from HDR images. In Section 4.2, we apply deconvolution methods. In Section 4.3 we describe our new method for removing glare, compare it to deconvolution, and discuss the results. Finally, in Section 7 we consider the applications and limits of our methods, and we conclude with future work.

2 Related Work

Measuring glare. Veiling glare is a commonly-acknowledged problem in photography. The standard method for glare measurement involves photographing a central black target surrounded by a large uniform bright illuminant [Matsuda and Nitoh 1972; Kuwabara 1953]. ISO standard 9358[1994] defines the ratio of luminance in the center of the target to the luminance of the illuminant as the veiling glare index (VGI). The standard also defines the glare spread function (GSF), which describes the amount of glare created by a small bright spot as a function of the distance from the center of the spot. Directly relevant to our work, McCann and Rizzi[2007] recently measured glare in multiexposure HDR imaging, and found

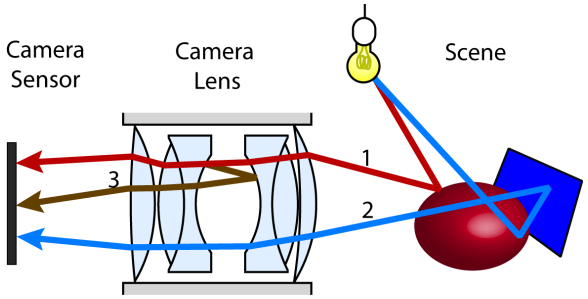

Figure 2: *The method of direct/global separation by Nayar et al.[2006] seeks to separate light rays that bounce directly from the scene to the camera (ray 1) from global illumination rays that bounce multiple times in the scene (ray 2). Analogously, we seek to separate rays 1 and 2, which are both direct rays in the camera lens, from glare-causing reflected and scattered rays (ray 3) which constitute global light transport in the camera lens.*

that the dynamic range of the resulting HDR capture was often limited by veiling glare, not by the range of exposures available on a camera.

Removing glare. Most methods for reducing veiling glare focus on improvements in the optical elements of the system. Better lens coatings, for example, greatly reduce the reflections from lens surfaces. A recent example of more radical methods can be found in the work of Boynton et al.[2003], who constructed a fluid-filled camera in order to reduce reflections from the lens surfaces.

Several computational techniques for removing glare have also been proposed. In medical imaging, deconvolution methods have been used to remove veiling glare computationally, for example in X-ray imaging [Seibert et al. 1985; Faulkner et al. 1989]. In astronomy, deconvolution methods are used for removing atmospheric distortion and scattering, as well as for other image enhancement tasks. Starck et al.[2002] present a review of both the theory of deconvolution and the techniques used to solve deconvolution problems in astronomy.

Glare removal in HDR photography is discussed by Reinhard et al.[2006], where they describe a method for estimating a global GSF for a scene based on fitting a radially-symmetric polynomial to the falloff of light around bright pixels. The method is a form of blind deconvolution, and hence is similar to the deconvolution method we describe.

All of the above computational methods post-process an image that already contains glare. In HDR scenes with significant veiling glare these methods will perform poorly, as signal-to-noise ratios are low in the presence of strong glare. To overcome this, glare must be physically removed before it is recorded.

Separating local and global light transport. Recent work by Nayar et al.[2006] on separating direct and global illumination using structured illumination or occlusion masks provide an approach to overcome these capture limitations. Their method was originally used to create novel images showing the two types of light transport in real-world scenes. Their method decomposes the transport into a direct component image, resulting from light that reflects directly from the scene to the camera, and an indirect component image, resulting from all multiple-bounce events in the scene. By considering the entire scene as a "virtual light source" and the internal camera optics as the "scene", veiling glare can be seen as the result of the indirect light transport in the scene, and a variant of this method can be applied to separate out glare before capture.

Methods have also been developed to measure the full 4D [Sen et al.

2005] or 8D [Garg et al. 2006] transport tensor between the light source and sensor, allowing synthetic relighting of a captured scene in the 4D case, or the ability to change the viewpoint in the 8D case. While these methods can be used to characterize glare, they do not provide an avenue for separating out glare before it is recorded on the camera image sensor.

3 Characterizing Glare

Before describing our methods to remove veiling glare, let us first characterize it, paying particular attention to its frequency behavior, variance across the camera's field of view, and magnitude. We constructed a testbed, consisting of a rotating camera and a white LED aimed at the center of rotation. HDR images were captured of the point light source as the camera was rotated over a wide angle. We used a standard multiexposure method [Debevec and Malik 1997] to capture a sequence of photographs that were then composited together. Only the exposure time of the camera was varied, because the lens flare changes with the camera aperture setting (f/stop).

We used three cameras for our measurements, to characterize the glare for a wide range of quality and price. The first was the Canon EOS 20D, a high-quality 8-megapixel digital single-lens-reflex (DSLR) camera. Like all SLR cameras, it can use many different lenses, each of which will exhibit a different amount of glare. The sensor on the 20D allows for 12 bits per pixel of precision in RAW mode. The second test camera was the Canon PowerShot A640, a 10-megapixel point-and-shoot camera. The A640 only outputs JPEG files, so the camera provides 8bpp of precision. Finally, we used a single camera tile from the Stanford Multi-Camera Array [Wilburn et al. 2005], roughly the quality of a commercial webcam: a resolution of 640x480 pixels and 8bpp of precision, connected to a simple fixed lens with no aperture or zoom adjustment.

Figure 3 shows representative samples of a sweep past the LED for all three cameras. The 20D was equipped with a Canon 28-135 mm IS USM lens, set to an aperture of f/8 and a focal length of 28 mm, and the A640 was set to an aperture of f/8 and moderate zoom. The HDR capture process for each set used the full exposure time range of each camera, from 30 seconds to 1/8000th of a second for the Canon 20D, in order to capture the full dynamic range in the GSF. Due to the limited exposure adjustment on the camera array tile, the dynamic range in its captures is much less than that of the 20D or the A640. Figure 4 shows a log-space surface plot for one flare for each camera. The central lobe of glare falls off rapidly, and the scene-wide diffuse veiling glare is very dim. For the Canon 20D, the uniform glare is 20 binary orders of magnitude (or 20 f-stops) dimmer than the point source. Bright images of the hexagonal lens aperture are also visible, at much higher brightness than veiling glare. Diffraction rays can also be seen, caused by the blades of the aperture diaphragm, also with a hexagonal symmetry. The A640 has similar artifacts, though notably it also has a 3x3 grid pattern of ghosts surrounding the point source. The number of ghosts and other artifacts increases with the number of surfaces in a lens system, so the 20D glare images have many more ghosts than the A640 glare images. The camera tile has only a few lens surfaces, and a fixed circular aperture, so its lens flare consists of mainly the central scattering from the inexpensive lens elements.

Based on our measurements, it is clear that glare has both high-frequency and low-frequency components. While aperture ghosts and diffraction rays quickly blur out for an extended source and simply contribute to overall veiling glare, the steep falloff in glare magnitude around a bright region evident in the figures gives glare a high frequency component near sharp luminance edges. However, in regions farther away from bright sources, glare is smooth. The magnitude of the glare 'floor' depends on the amount of light present in the scene. Based on simulations, the Canon 20D can record nearly 20 stops of dynamic range using HDR imaging if only a point light source is present. If half of the field of view is covered by an extended source, then only 9 stops of dynamic range can be recorded by the 20D; the A640 would only record 7 stops. Since most of the glare is low-frequency, we will focus our efforts on removing this component of veiling glare.

(a) (b) (c)

Figure 3: *Tonemapped GSFs for three positions of a point light source for three cameras. Each column shows GSFs for one camera, depicted on the bottom row. (a) Canon EOS 20D, (b) Canon PowerShot A640, (c) Camera tile from the Stanford Multi-Camera Array. Note that the glare patterns are not shift-invariant.*

4 Removing Glare

4.1 Notation

Before discussing our two methods for removing veiling glare, it will be helpful to introduce some notation. We wish to reconstruct an $(n \times m)$ glare-free image $S(u,v)$, based on a recorded $(n \times m)$ image $R(x,y)$ that contains glare. The glare-free image is the image that would be created by an ideal lens system otherwise identical to the one the recording camera has, except that it is free of veiling glare. In other words, a point light source in focus is recorded as a single point in this ideal camera; all other lens system effects, such as depth-of-field defocus, vignetting, and any geometric or chromatic aberrations, are still present in S.

The process of glare formation can be described as a linear transform from S to R. The transform is composed of the 2D glare spread functions for each pixel in S, resulting in a 4D tensor $\mathbf{L}(x,y,u,v)$. Each 2D plane of $\mathbf{L}$ with constant u and v, $L_{u,v}(x,y)$, is a GSF for pixel (u,v) in the glare free image S. If $\mathbf{L}$ is sliced into 2D planes of constant x and y instead, each 2D plane $\mathbf{L}_{\mathbf{x},\mathbf{y}}(u,v)$ describes how much glare pixel (x,y) in R receives from each pixel in S. We will refer to this 2D function as the glare transfer function (GTF) for pixel (x,y).

For notational simplicity, instead of dealing with 2D images and a 4D tensor, we convert the images to $(nm \times 1)$ vectors $\mathbf{s}(u)$ and $\mathbf{r}(x)$,

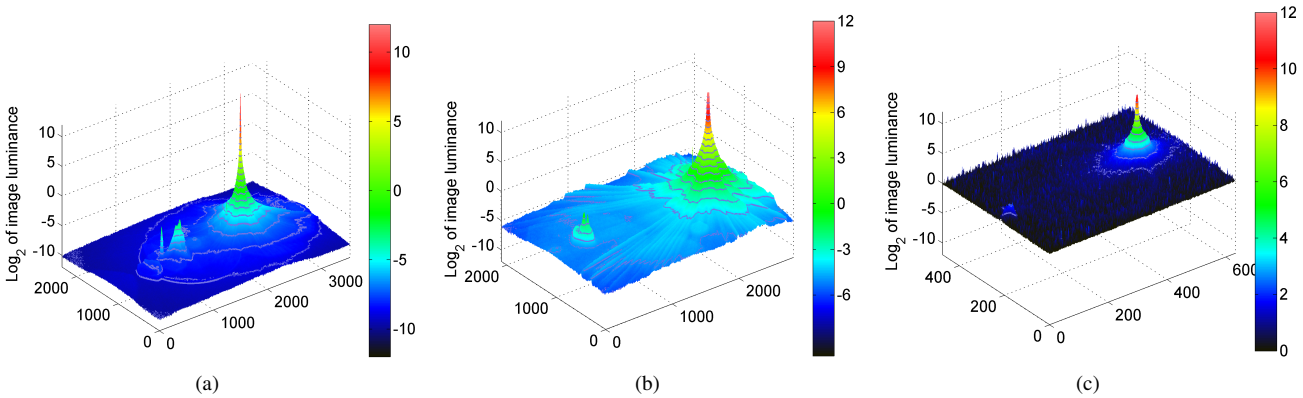

Figure 4: *Plot of glare magnitude for glare spread functions (GSFs) depicted on the second row of Figure 3. Relative luminance is plotted on a log scale, with blue contour lines drawn for every f-stop. The plots show the full field of view of each camera. The Canon 20D (a) has the least overall glare, about 4 stops less than the Canon A640 (b). The noise floor of the camera array tile (c) is high enough to obscure most of its glare, and makes it unsuitable for HDR captures of more than 10 f-stops.*

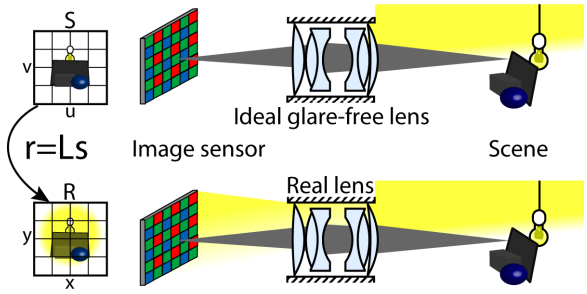

Figure 5: *Conceptual arrangement for glare removal. We attempt to reconstruct* **s***, an image that would be formed by an ideal glare-free camera, which is related to the recorded image* **r** *by* $\mathbf{r} = L\mathbf{s}$.

and the transform to a $(nm \times nm)$ matrix $L(x,u)$. Each entry of L describes how much light pixel u in $\mathbf{s}$ contributes to pixel x in $\mathbf{r}$. Each row $\mathbf{L_x}$ of L describes the glare transfer function $\mathbf{gtf_x}(u)$ for pixel x in $\mathbf{r}$, and each column $\mathbf{L_u}$ of L describes the glare spread function $\mathbf{gsf_u}(x)$ for pixel u in $\mathbf{s}$. The fundamental equation of glare formation is then a straightforward matrix multiplication:

$$\mathbf{r} = L\mathbf{s} \tag{1}$$

Figure 5 illustrates this arrangement. On top is the ideal lens system that produces $\mathbf{s}$, and on the bottom is the real camera that produces image $\mathbf{r}$.

We seek to invert Equation 1 first through a deconvolution technique, and then through our novel occlusion mask method.

4.2 Deconvolution method

If the glare spread functions $\mathbf{gsf_u}$ that make up L are all shifted versions of each other:

$$\mathbf{gsf_u}(x) = \mathbf{gsf_v}(x+(u-v)) \quad \forall u,v \in 1\ldots nm \tag{2}$$

then the GSF is shift-invariant, and we only need one representative glare spread function, $\widehat{\mathbf{gsf}}$ to describe all of L. The formation of the recorded image $\mathbf{r}$ can be described as a 2D convolution:

$$\mathbf{r} = \mathbf{s} * \widehat{\mathbf{gsf}} \tag{3}$$

This assumption is generally unjustified. As shown in Figure 3, the GSF varies across the scene. However, if we assume the variance is negligible, we can use deconvolution to solve for $\mathbf{s}$ given $\mathbf{r}$ and $\widehat{\mathbf{gsf}}$ [Starck et al. 2002]. Specifically, we can model our GSF as the desired direct component $\delta_\mathbf{x}$, a delta function, plus some low-valued low frequency component, $\mathbf{b}$, representing the undesired glare. We then use a single step gradient descent deconvolution method, which estimates $\mathbf{s}$ by convolving $\mathbf{r}$ with $\delta_\mathbf{x} - \mathbf{b}$:

$$\begin{aligned}
(\delta_\mathbf{x} - \mathbf{b}) * \mathbf{r} &= (\delta_\mathbf{x} - \mathbf{b}) * \widehat{\mathbf{gsf}} * \mathbf{s} && (4)\\
&= (\delta_\mathbf{x} - \mathbf{b}) * (\delta_\mathbf{x} + \mathbf{b}) * \mathbf{s} && (5)\\
&= (\delta_\mathbf{x} - \mathbf{b} * \mathbf{b}) * \mathbf{s} && (6)\\
&\approx \mathbf{s} && (7)
\end{aligned}$$

The peak value of $\mathbf{b}$ will be several orders of magnitude less than the value of the direct component, so the second order term $\mathbf{b} * \mathbf{b}$ is several orders of magnitude smaller again. Given this, the problem is extremely well conditioned, and so a single step gradient descent deconvolution will work well.

To demonstrate this idea, we estimated $\mathbf{b}$ by hand for the A640 as a combination of Gaussians of different widths, and then performed deconvolution to produce the results shown in Figure 6.

As argued above, there are two main problems with this method. First, we can see that the values in glare-dominated regions have become noisy due to the quantization limit. There simply is not enough real scene information in the captured image for any post-processing method to recover. Second, we also see some residual glare of different colors in various parts of the image. There is no one $\widehat{\mathbf{gsf}}$ that will work for the entire image. The shape, size, and even color characteristics of the GSF vary across the field of view, and even a hand-tuned $\widehat{\mathbf{gsf}}$ constructed to produce the best looking output does not perform well.

4.3 Structured occlusion mask method

As shown above, quantization in the presence of significant glare results in poor glare removal. Therefore, we propose placing a high-frequency occlusion mask between the camera and the scene, physically limiting the amount of glare created in the camera. If we use a mask with an average occlusion factor of α, and we assume glare consists of only low spatial frequencies, we then would expect the glare recorded at all pixels to be reduced by a factor of α as well.

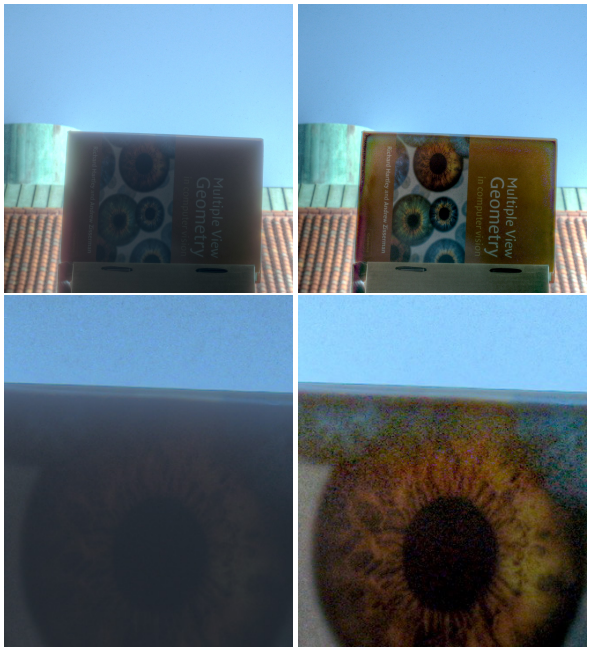

Figure 6: *A deconvolution experiment using the Canon A640. The scene is a book in a dark room, in front of a bright window showing the roof of a neighboring building and some blue sky. The top row shows a regular HDR capture of a scene, and the result from using deconvolution to remove glare. The bottom row shows an inset of the top of the book, demonstrating how the quantization noise is amplified by the deconvolution. All images are tonemapped.*

This global transport low-frequency condition was the key insight in the work of Nayar et al.[2006]. While in our case, the condition only holds approximately, it still allows us to estimate the expected improvement in reconstruction. For simplicity of construction and analysis, we limit ourselves to a binary mask

In an HDR image taken using a standard multi-exposure capture, the precision of a pixel is at most equal to the precision of the camera image sensor, since only well-exposed pixels from each exposure are used for the final HDR image. As an example, consider a pixel near a bright scene feature in a normal HDR capture. The glare g at that pixel could be 15 times greater than the true image value d, and the HDR image will be record the sum $g+d \approx 16d$ with the precision of the image sensor. With an 8-bit sensor, only 4 bits are used to record d, the value we seek to recover. Therefore, any reconstruction will be poor. With a 90% opaque mask in place, however ($\alpha = 0.1$), g is reduced substantially: $\alpha g + d \approx 2.5d$. d will be recorded with 7 bits of accuracy. This improvement in SNR allows for much less error in the reconstruction. Figure 7 illustrates the effect.

This creates a basic tradeoff between reconstruction SNR and capture time. With a small α, we minimize glare, but many captures are needed to record all scene content through the gaps in the mask - at the limit, one is reduced to scanning a pixel-sized hole across the scene. A large α allows for a quick capture, at the cost of increased noise. This tradeoff also suggests that employing sinusoidal mask patterns, such as those used by Nayar et al.[2006], would allow only slight SNR improvements. Such a pattern would have $\alpha = 0.5$, so the improvement in reconstruction accuracy from a sinusoidal occlusion pattern would be dwarfed by the increase in noise.

For a given α, at least $\frac{1}{\alpha}$ captures are needed. The mask must be translated between each capture, to ensure all scene regions are

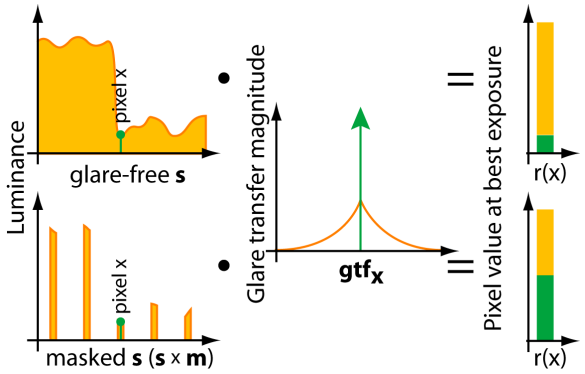

Figure 7: *A 1-D illustration of the SNR improvement due to an occlusion mask. On the top, the pixel that images the green spot in the scene receives the direct light from the green spot (the peak in the glare transfer function) as well as all the glare from the rest of the scene, resulting in a sum that is mostly glare (the yellow top part of the bar on the right). In this sum, few bits are devoted to the direct light. On the bottom, an occlusion mask with $\alpha = \frac{1}{6}$ reduces the glare at the pixel to be roughly equal to the direct component (equal sizes of bottom green and top yellow bars), allowing for a longer exposure and improving quantization.*

Figure 8: *Periodic artifacts in direct/global separation. The pattern is aligned to the mask orientation and strongest near large brightness gradients, but is visible across most of the scene.*

imaged through a hole. Doing this gives us at least one recording of $d + \alpha g$ at each pixel, and many recordings of roughly αg.

If the assumption of Nayar et al.[2006] were to hold for glare, the occluded pixel values would all be exactly αg, and we could simply subtract the two values recorded for each pixel to obtain d. However, unlike global light transport in typical scenes, the GTF has significant high-frequency components, as can be seen in Figure 4. Therefore, glare recorded for an occluded pixel varies across captures. Attempting a subtraction leads to significant artifacts, as shown in Figure 8.

We avoid these artifacts by considering a single captured HDR image, instead of the stack of values captured per pixel. Within a single capture, the glare varies relatively smoothly, and can be estimated for unoccluded pixels based on nearby occluded ones. Once we have such a glare estimate per pixel, we can subtract it out. Since the estimate is low-frequency, a residual high-frequency component of glare remains. We then composite the individual glare-subtracted captures together to form a complete image of the scene.

As the size of the mask holes is shrunk relative to the glare spread function, the glare estimates become more accurate. Again, a tradeoff exists between the number of required captures and the quality of the results. Nayar et al.[2006] use a similar interpolation to do

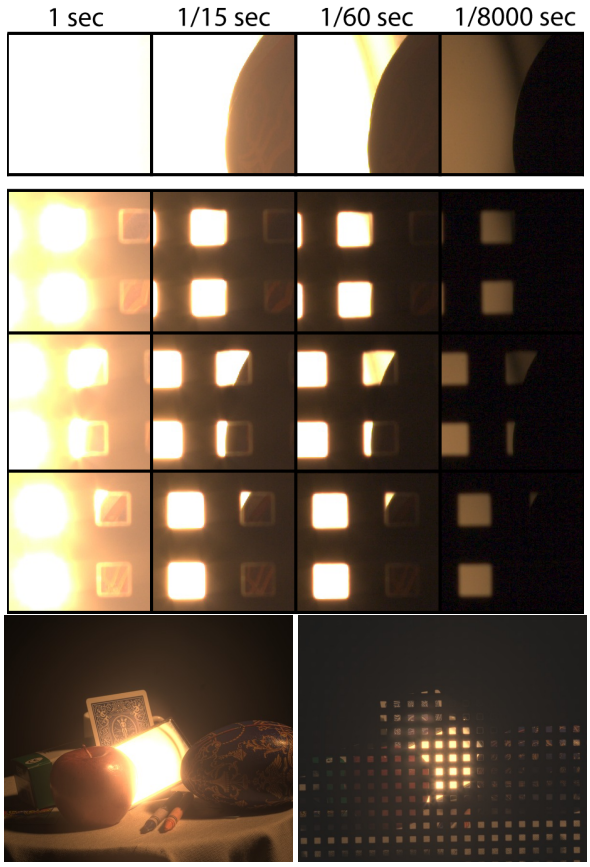

Figure 9: *A close-up of a sequence of captures. Images in the same column all have the same exposure time; four exposure times out of the total of 8 for this capture are shown. The first row has raw images for the unmasked scene. The next three rows show raw images for three different grid positions, out of the total of 36 for this experiment. These are the source images for compositing HDR images. Two composited HDR images are shown on the last row, corresponding to the non-masked capture and to the topmost of the masked captures.*

direct/global decompositions with a single photograph, with the difference being that they spatially interpolate the direct component as well as the global component. Our method requires knowledge of exactly where the mask is, so we must first register each image to a projectively warped mask pattern. In the following section we will describe how we acquire our images, and then the details of the registration and interpolation.

5 Implementation

5.1 Acquisition

In theory, any high-frequency pattern can be employed as an occlusion mask. However, making a large mask that partially attenuates light, instead of merely blocking it, is difficult. The mask must not introduce any additional glare, shine, or scattering, so a translucent mask would have to be constructed from high-quality optical material, and kept clean of dust at all times, which is difficult to do with a large mask placed in a scene. We instead employ a binary mask: a black cardboard sheet with holes cut out of it with a laser cutter.

The minimum number of captures needed to image every scene point through a hole is $\frac{1}{\alpha}$. To reach this minimum, each capture must not re-image regions already imaged by other captures. An easy way to meet this requirement is to use a simple periodic mask pattern. We chose a regular grid of square holes, which allows us to simply translate the mask through a regular grid of positions to ensure full coverage. While the sampling characteristics of such a mask are not optimal, the mask is easy to construct and to use.

However, even with our simple mask, more than the minimum number of captures are required. This is due to defocus blur of the mask, and mask edge effects. Figure 9 shows close-ups of raw captures. For the longest exposure captures, the hole edges clearly exhibit both problems - a bright blurry halo surrounds the true scene content. These halos are due to grazing angle reflection from the mask edges, and are present even for very diffuse mask materials. The halos are out of focus because the camera is focused on the scene, behind the mask. We do not try to compensate for these halos, and instead simply discard the entire border region for each hole. Since this reduces the number of useful pixels per capture, the number of captures must be increased. Based on our experiments, a mask that would require a 3x3 set of HDR captures instead needs a 6x6 grid of captures to ensure good coverage. While thinner masks may reduce the halo effects, for most camera apertures defocus is unavoidable, since the mask cannot be placed in the middle of the scene.

5.2 Reconstruction

After acquisition, we have more than $\frac{1}{\alpha}$ HDR images, each imaging the scene with a different mask position, and each containing roughly α times the glare of the mask-free HDR image. The edge effects discussed above are present, and must first be removed.

For each capture, we fit a perspective-warped planar grid (8 unknowns) using an iterative grid growth technique to find the planar homology. An initial 3×3 grid estimate is placed in the scene. Then each estimate grid hole center is moved to the centroid of its local luminance distribution, a new least-squares grid homology is found based on the new mask hole centers, and the algorithm is iterated until the error is sufficiently low. Then the grid is expanded to 5×5, and the iteration is repeated, the grid is again expanded, and so on. If the homology looks incorrect, the grid is relocated to a different region of the captured image and the iteration is restarted. We assume a homology is incorrect if its coefficients vary significantly from the initial estimate. After the homology has been found, we use it to construct two post-processing masks, $\mathbf{m}_\phi^+$ and $\mathbf{m}_\phi^-$ for each occluding mask position ϕ. The first discards everything but the unblocked scene regions in the occluding mask holes, and the second discards everything but the central areas of the blocked scene regions.

The reconstruction proceeds on a per-mask-position basis. For capture $\mathbf{r}_\phi$, consider the digitally masked image $\hat{\mathbf{g}_\phi} = \mathbf{m}_\phi^- \times \mathbf{r}_\phi$, where $\times$ indicates elementwise (Hadamard) multiplication. This image only contains global glare values for occluded regions. We then interpolate $\hat{\mathbf{g}_\phi}$ across the unoccluded regions using a weighted Gaussian blur with standard deviation of one third of the mask pattern period. If $\mathbf{f}$ is the Gaussian blur kernel, then

$$\mathbf{g}_\phi = \frac{\mathbf{f} * (\mathbf{m}_\phi^- \times \mathbf{r}_\phi)}{\mathbf{f} * \mathbf{m}_\phi^-} \tag{8}$$

where the division is elementwise. This gives us a low-frequency glare estimate $\mathbf{g}_\phi$ for the entire capture.

Next, we subtract the estimated glare from the original captured image, and discard the regions outside the mask holes to create a

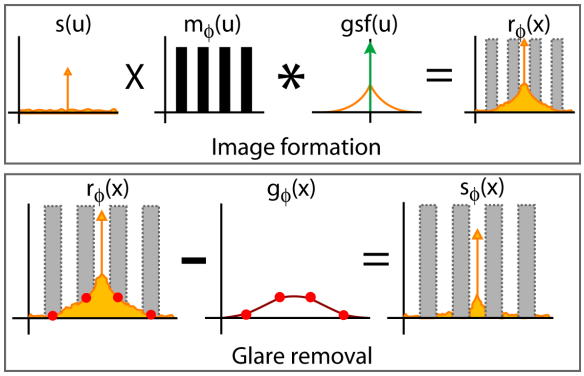

Figure 10: *Illustration of glare removal for a 1D scene with a point light source and a lens with a shift-invariant GSF. On top, image formation with an occlusion mask: The ideal image is multiplied elementwise by the mask and then convolved by the GSF of the lens to form one recorded image* $\mathbf{r}_\phi$*. The bars are regions occluded by the mask. On bottom, the glare interpolation procedure: The occluded points in* $\mathbf{r}_\phi$ *marked with red circles are used to interpolate a glare estimate* $\mathbf{g}_\phi$*, which is subtracted from the unoccluded regions of* $\mathbf{r}_\phi$ *to create glare-free estimate* $\mathbf{s}_\phi$*. Not shown is the final compositing step to combine all* $\mathbf{s}_\phi$ *into a single glare-free* $\mathbf{s}$*.*

glare-free estimate for this capture, $\mathbf{s}_\phi$:

$$\mathbf{s}_\phi = \mathbf{m}_\phi^+ \times (\mathbf{r}_\phi - \mathbf{g}_\phi) \tag{9}$$

The interpolated estimate is generally dimmer than the true glare, and therefore the subtraction is conservative. This condition may fail at strong brightness edges, where the glare estimate can exceed the true glare value on the dark side of the edge.

After repeating this process for all the captures, we have a full set of glare-removed images, each containing a section of the final image. We then blend these images together with a weighted average to form the final result, $\mathbf{s}$. For each image, the weights fall off as a Gaussian from the center of each mask hole. The weights are set to zero for all pixels that are outside the central hole regions. The final reconstruction algorithm is then:

1. For each captured HDR image $\mathbf{r}_\phi$:
   (a) Estimate the pose of the occlusion mask in the capture, and construct $\mathbf{m}_\phi^+$ and $\mathbf{m}_\phi^-$.
   (b) Create $\mathbf{g}_\phi$ using Equation 8
   (c) Create $\mathbf{s}_\phi$ using Equation 9
2. Composite all $\mathbf{s}_\phi$ together to form glare-free reconstruction $\mathbf{s}$.

Figure 10 demonstrates the sequence for a simple 1-D scene with a single point light source, and Figure 11 illustrates the reconstruction process for an actual captured image.

6 Experimental Demonstration

For our experiments, we constructed a mask with a 30x30 grid of square holes 4 mm in width and a period of 1 cm. This creates an occlusion factor α of 0.16. The grid was made from matte black cardboard, which is thin (0.2 mm thick) to minimize edge effects, but completely opaque. The capture system consisted of a motorized Lego tripod mount capable of translating the mask in two dimensions and a laptop controlling the mount and the cameras used in the experiment. Figure 12 shows a typical arrangement of the scene, the mask, and the camera. With holes 4mm on the side, the

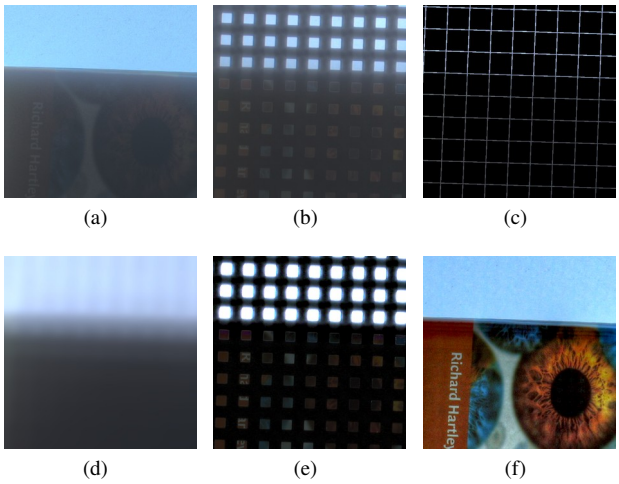

Figure 11: *A closeup of a tonemapped sequence of HDR images formed in interpolating glare removal. (a) Unoccluded original. (b) Single occluded capture* $\mathbf{r}_\phi$*. The occluded regions are extracted to produce (c), the glare-only image* $\hat{\mathbf{g}}_\phi$*. This image is interpolated to form (d), the glare estimate* $\mathbf{g}_\phi$*.* $\mathbf{g}_\phi$ *is then subtracted from (b) to produce (e), the glare-free estimate* $\mathbf{s}_\phi$*. Compositing all* $\mathbf{s}_\phi$ *together forms (f), the complete HDR glare-free image* $\mathbf{s}$*.*

mask creates no significant diffraction effects, based on both theoretical calculations and simple experiments.

We again used the Canon 20D DSLR camera and the Canon A640 point-and-shoot for glare removal. We did not use the camera tile, because of its limited dynamic range. 36 mask positions in a 6x6 grid were sufficient to capture every pixel, and a complete capture took between 30 and 60 minutes, depending on the camera and the number of exposures needed. To tonemap our results, we used the gradient domain method of Fattal et al.[2002].

Figure 1 is a scene of a number of objects in front of a strongly illuminated white background, captured with the Canon 20D. The background brightness was measured to be 720 cd/m^2, while the foreground objects were below the 1 cd/m^2 lower limit of the light meter used, giving a lower bound of 720:1 to scene dynamic range. In the original HDR capture, the details in the foreground objects are nearly completely obscured. Due to the SNR improvement of a factor of $1/\alpha \approx 6$ inherent in our method, the uniformly veiled regions are recovered with full detail.

At strong luminance edges where high-frequency glare is most apparent, we fail to recover scene content. Therefore, regions such as the top of the mirror or the blue egg become black after glare removal. Because of these artifacts, presenting whole-scene dynamic range figures would be inappropriate. However, quantitative comparisons of the HDR images between specific regions away from strong edges can be made. Measuring mean brightness for the lens diagram in the background, and the central area of the blue egg in the foreground, a dynamic range of 560:1 was measured before glare removal. After glare removal, the dynamic range improved to 22400:1. Similarly comparing the lens diagram to the front of the apple, the dynamic range went from 505:1 before glare removal to 8850:1 after. Comparing these regions shows how the glare almost completely obscured the 2:1 real brightness difference between the apple and the egg. The ratios after glare removal are consistent with the light meter lower bound.

Figure 13 shows a real-world scenario of a scene strongly veiled by glare from a bright window. In the original HDR capture, the

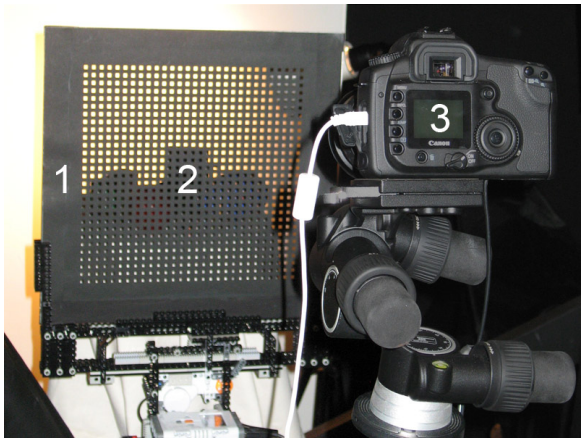

Figure 12: *A typical arrangement of the occlusion mask (1), the scene behind it (2), and the camera (3). The mask is as close to the scene as possible to minimize defocus effects, and is mounted on a gantry which translates the mask between each HDR capture. Over the course of the experiment, the mask is translated by a total of 1 cm in X and Y.*

book is obscured by glare. At the top of the book, the glare is the color of the sky, and at the bottom, it is the color of the building across the street. The reconstruction has neither glare component remaining - the glow at the bottom right is light reflecting from an off-scene object back onto the book, a subtle effect not visible before glare removal. Along the book edges, some ringing can be seen, due to high-frequency glare components left in the image. Compared to the deconvolution approach shown in Figure 6, our method shows far less noise in glare-obscured regions. Comparing the mean brightness of a sky patch to the title on the book cover, we find a dynamic range of 206:1 before, and 375:1 after glare removal.

Finally, Figure 14 shows a miniature scene of a snowboarder at night, back-lit by a mock full moon. The scene is very dark, resulting in high noise even for the Canon 20D and a 30-second exposure. However, the result clearly shows more detail in glare-obscured regions of the snowboarder, and in the black fabric in the background. The main artifacts are the halos resulting from remaining high-frequency glare, mainly seen inside the silhouette of the snowboarder. Here, comparing the mean brightness of the center of the moon to the chest of the snowboarder, we find a dynamic range of 435:1 before, and 2850:1 after glare removal.

7 Conclusion and Future Work

We have developed a new method for removing veiling glare from HDR images, using multiple captures with a high-frequency occlusion mask to estimate and remove glare. Glare is reduced before image formation, improving SNR and providing high-quality reconstructions. Both SNR and the estimation accuracy improve as the mask occlusion factor α is decreased, leading to a trade-off between quality and acquisition time. We have also shown that pure post-processing such as deconvolution results in high quantization noise in strongly veiled regions, due to insufficient precision in the captured images.

The most significant limitation of our method is that it requires a large number of photographs to record a scene, so we can only apply it to static scenes. A complete capture with our setup takes from 30 minutes to an hour. The other major limitation is that the mask must be nearly in focus to limit the mixture pixel region. This mandates either using a very small aperture, or placing the mask near the scene rather than on the camera lens like a filter. In turn, this

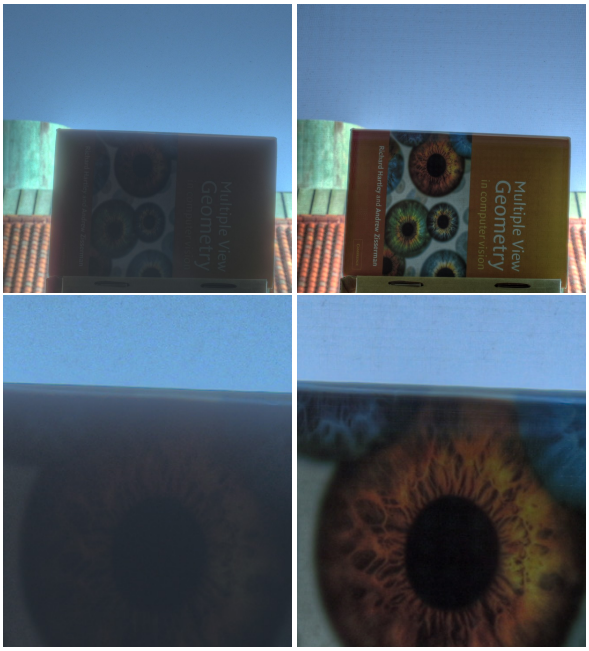

Figure 13: *Tonemapped images of a book in a dark room, against a bright window showing a neighboring building and blue sky, captured with the Canon A640. At the top left is the unmasked HDR capture, and at the top right is our glare-free result. At the bottom is a closeup of corresponding regions from each capture. Compare to the deconvolution result in Figure 6.*

restricts our technique to laboratory or studio settings.

If a small aperture is acceptable for the application, it is possible to mount the occluding mask to the front of the camera lens. At an aperture of f/22 and no zoom on the Canon 20D with the Canon 28-135 IS USM lens, the defocus blur of a mask attached to the front of the camera lens is roughly 30 pixels wide. Therefore, a mask mounted on the front of the lens with holes 90 pixels wide as seen by the camera would require about three times the acquisitions that an in-focus grid would. This suggests using a large α for reasonable capture times. For a security camera application, for example, $\alpha = 0.5$ and $p = 90$ pixels and a defocus grid blur of 30 pixels would require roughly 6 captures per glare-reduced image. Since security cameras are often HDR imaging systems on their own, with a fast-moving mask one could generate glare-free security camera video at 5 fps. Such a mask could be created by using a LCD panel in front of the camera lens, like the one used by Nayar et al.[2003] for adaptive dynamic range imaging.

Our method could also be adapted for capturing glare-free environment maps using a gazing ball. Our mask could be painted onto the ball, and the ball then rotated between captures. The only additional requirement would be accounting for the ambient illumination falling on the dark grid regions.

The occlusion mask used in our captures was a rectilinear grid, selected because it was easy to construct and use. Sampling theory tells us that such a grid pattern has poor sampling characteristics. We therefore hope to investigate staggered or pseudo-random grids. One could, for example, construct stratified stochastic grids which, despite their randomness, retain some limit on the number of captures required. A non-uniform mask makes registration slightly more challenging, but by no means impossible. Finally, a thinner occlusion mask would reduce the edge halo artifacts we observed, and further improve the accuracy of our glare-free reconstruction.

Figure 14: *Tonemapped results for a mock scene of a snowboarder against a 'moon' at night. Captured with the 20D, the scene is at the edge of the camera's sensitivity, with significant noise in both the original on top and the glare-free result on bottom . More detail is visible in the reconstruction, as can be seen in the inset. High-frequency glare remains visible as a band of light just inside the silhouette of the snowboarder.*

8 Acknowledgements

This work was funded in part by National Science Foundation grant CCF-0540872, and is partly based upon work supported under a Reed-Hodgson Stanford Graduate Fellowship. We would also like to thank John McCann and Alessandro Rizzi for fruitful discussion and advice.

References

BOYNTON, P. A., AND KELLEY, E. F. 2003. Liquid-filled camera for the measurement of high-contrast images. SPIE, D. G. Hopper, Ed., vol. 5080, 370–378.

DEBEVEC, P. E., AND MALIK, J. 1997. Recovering high dynamic range radiance maps from photographs. In *Proceedings of ACM SIGGRAPH 1997*, ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 369–378.

FATTAL, R., LISCHINSKI, D., AND WERMAN, M. 2002. Gradient domain high dynamic range compression. In *ACM Transactions on Graphics*, ACM Press, New York, NY, USA, 249–256.

FAULKNER, K., KOTRE, C., AND LOUKA, M. 1989. Veiling glare deconvolution of images produced by x-ray imageintensifiers. In *Proceedings of the Third International Conference on Image Processing and its Applications*, 669–673.

GARG, G., TALVALA, E.-V., LEVOY, M., AND LENSCH, H. P. A. 2006. Symmetric photography: Exploiting data-sparseness in reflectance fields. In *Proceedings of the Eurographics Symposium on Rendering*, 251.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. 1994. *ISO 9358:1994: Optics and optical instruments – Veiling glare of image forming systems – Definitions and methods of measurement*. International Organization for Standardization, Geneva, Switzerland.

KUWABARA, G. 1953. On the flare of lenses. *Journal of the Optical Society of America 43*, 1, 53.

MANN, S., AND PICARD, R. W. 1995. Being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures. In *IS&T 46th annual conference*, 422–428.

MATSUDA, S., AND NITOH, T. 1972. Flare as applied to photographic lenses. *Applied Optics 11*, 8, 1850.

MCCANN, J. J., AND RIZZI, A. 2007. Veiling glare: the dynamic range limit of HDR images. In *Human Vision and Electronic Imaging XII*, SPIE, vol. 6492.

NAYAR, S. K., AND BRANZOI, V. 2003. Adaptive dynamic range imaging: Optical control of pixel exposures over space and time. In *IEEE International Conference on Computer Vision*, 1168–1175.

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. In *ACM Transactions on Graphics*, ACM Press, New York, NY, USA, 935–944.

REINHARD, E., WARD, G., PATTANAIK, S., AND DEBEVEC, P. 2006. *High Dynamic Range Imaging - Acquisition, Display and Image-based Lighting*. Morgan Kaufman Publishers, 500 Sansome Street, Suite 400, San Francisco, CA 94111.

SEIBERT, J. A., NALCIOGLU, O., AND ROECK, W. 1985. Removal of image intensifier veiling glare by mathematical deconvolution techniques. *Medical physics 12*, 3, 281–288.

SEN, P., CHEN, B., GARG, G., MARSCHNER, S. R., HOROWITZ, M., LEVOY, M., AND LENSCH, H. P. A. 2005. Dual photography. In *ACM Transactions on Graphics*, ACM Press, New York, NY, USA, 745–755.

STARCK, J., PANTIN, E., AND MURTAGH, F. 2002. Deconvolution in astronomy: A review. *Publications of the Astronomical Society of the Pacific 114*, 1051–1069.

WILBURN, B., JOSHI, N., VAISH, V., TALVALA, E.-V., ANTUNEZ, E., BARTH, A., ADAMS, A., HOROWITZ, M., AND LEVOY, M. 2005. High performance imaging using large camera arrays. In *ACM Transactions on Graphics*, ACM Press, New York, NY, USA, 765–776.