

Removing Photography Artifacts using Gradient Projection and Flash-Exposure Sampling

Amit Agrawal

Ramesh Raskar

Shree K. Nayar[†]

Yuanzhen Li

Mitsubishi Electric Research Labs (MERL), Cambridge, MA*

[†]Columbia University

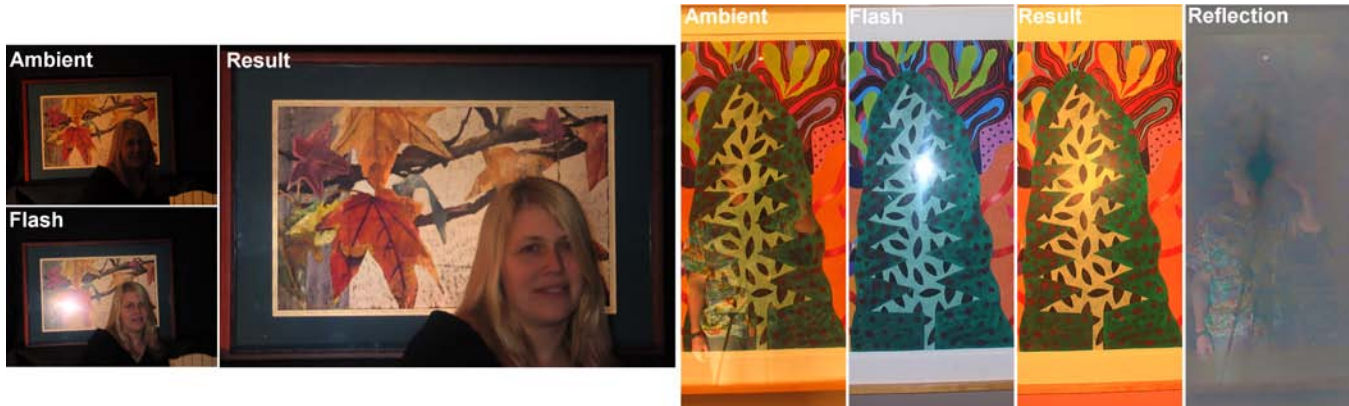


Figure 1: Undesirable artifacts in photography can be reduced by comparing image gradients at corresponding locations in a pair of flash and ambient images. (Left) Removing flash hot spot. Flash and ambient images of a museum scene, where the flash image reveals more of the scene but includes a strong highlight. We combine gradients in flash and ambient images to produce an enhanced flash image with the highlight removed. (Right) Removing self reflections. Flash and ambient images of a painting, where the ambient image includes annoying reflections of the photographer. The low-exposure flash image avoids reflections, but has a hot spot. We remove the reflections in the ambient image by removing the component of the ambient image gradients perpendicular to the flash image gradients. For visual verification, we show the computed reflection layer.

Abstract

Flash images are known to suffer from several problems: saturation of nearby objects, poor illumination of distant objects, reflections of objects strongly lit by the flash and strong highlights due to the reflection of flash itself by glossy surfaces. We propose to use a flash and no-flash (ambient) image pair to produce better flash images. We present a novel gradient projection scheme based on a gradient coherence model that allows removal of reflections and highlights from flash images. We also present a brightness-ratio based algorithm that allows us to compensate for the falloff in the flash image brightness due to depth. In several practical scenarios, the quality of flash/no-flash images may be limited in terms of dynamic range. In such cases, we advocate using several images taken under different flash intensities and exposures. We analyze the flash intensity-exposure space and propose a method for adaptively sampling this space so as to minimize the number of captured images for any given scene. We present several experimental results that demonstrate the ability of our algorithms to produce improved flash images.

Keywords: Flash, reflection removal, gradient projection, flash-exposure sampling, high dynamic range (HDR) imaging

*emails: aagraval@umd.edu (Currently at UMD), raskar@merl.com, nayar@cs.columbia.edu, yzli@mit.edu (Currently at MIT).
Web: <http://www.merl.com/people/raskar/photo/>

Copyright © 2005 by the Association for Computing Machinery, Inc.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions Dept, ACM Inc., fax +1 (212) 869-0481 or e-mail permissions@acm.org.

© 2005 ACM 0730-0301/05/0700-0828 \$5.00

1 Introduction

Flashes are often used to capture a good photograph of a scene under low-light conditions. However, flashes produce a variety of undesirable effects and artifacts. They tend to saturate nearby objects while failing to light up distant ones. Since the flash intensity falls with distance from the camera, flashes produce a tunnel effect, where brightness decreases quickly with depth. Furthermore, flashes are notorious for producing undesirable reflections. Often one sees the reflection of an object that lies outside the field of view of the camera but is strongly lit by the flash, and/or by a specular object within the field of view. One also sees strong highlights due to direct reflection of the flash itself by glossy objects in the scene. Previous work has shown that a flash and ambient image pair can be used to de-noise and enhance the ambient image and to achieve enhanced image fusion. Our goal is to use a flash and ambient image pair to produce a high quality flash image. Our main contributions can be summarized as follows:

- We present a gradient orientation coherence model that relates gradients in the flash and ambient images. This model seeks to capture the properties of image gradients that remain invariant under the change of lighting that takes place between a flash and an ambient image. Based on the coherence model, we propose a gradient projection method that removes the component of image gradients that are introduced by undesirable reflections.
- We show that the ratio of brightness in the flash and the ambient image represents the combined effects of depth and surface orientation. Using this ratio, we show how a flash image can be modified to reduce the fast brightness falloff due to increase in depth.
- In principle, only a single flash-ambient image pair is needed

to achieve all of the above enhancements. In practice, however, the flash or the ambient image may be of very low quality due to the limited dynamic range of the camera. We analyze the two-dimensional space of flash intensities and exposure settings. We show how, for a given scene, one can adaptively sample this space to minimize the number of images needed to obtain high quality flash and ambient images.

The methods we propose have the following limitations:

- Our approach requires a minimum of two images of the scene to be captured. This restricts its applicability to *static scenes*. This limitation is shared by many high dynamic range (HDR) capture methods as well as image based rendering methods.
- We cannot handle *co-located artifacts* in the ambient and the flash image, such as a hot spot in the flash image and reflections in the ambient image that appear at the same location. For the gradient projection scheme to work, we require that for every pixel either the flash or the ambient image be free of artifacts.
- Since our methods are based on the manipulation of image gradients, the enhanced images may include non-monotonic intensity mappings. In short, our goal is to produce images that are pleasing to the eye rather than ones that represent accurate radiometric measurements of the scene.

2 Related Work

Flash/no-flash Pairs: In recent work, flash images have been used to significantly enhance details and reduce noise in ambient images [Petschnigg et al. 2004; Eisemann and Durand 2004]. Eisemann and Durand also take colors from the flash image and achieve an enhanced image fusion of flash and ambient images. These previous papers have also discussed the removal of flash shadows, red-eye reduction and white balancing. Our work draws inspiration from the previous work, but our goals are different. We seek to reduce the strong artifacts produced by a flash and generate better flash images. In case of indoor scenes with shallow depth ranges, flash illuminates most of the scene adequately and hence the flash image can be used to enhance the quality of an ambient image. Here, we are interested in scenes that might include large depth variations and significant variations in ambient illumination. Although noise is not our main concern, it is worth noting that in such scenes noise in a low-exposure flash image can actually be higher than that in the ambient image. For example, distant points which are not well-lit by the flash may be captured with a lower SNR in a low-exposure flash image as compared to the ambient image. To enhance the ambient image, previous methods have used variants of the joint bilateral filter. Instead, we use gradient domain techniques which are more effective for detecting and reducing large artifacts such as reflections.

Gradients for Image Fusion: Our work is also related to the emerging space of imaging techniques that combine two or more images. For example, HDR is achieved by varying exposure [Mann and Picard 1995; Debevec and Malik 1997; Nayar and Mitsunaga 2000]. Our gradient domain techniques are related to those used for HDR compression [Fattal et al. 2002], image editing [Perez et al. 2003], image fusion for context enhancement [Raskar et al. 2004b], creating photomontages [Agarwala et al. 2004] and image matting [Sun et al. 2004]. Gradient direction matching has also been used in stereo matching and particle filtering [Lichtenauer et al. 2004]. Gradient-based invariants have been used for shadow removal [Finlayson et al. 2002] and face recognition [Chen et al. 2000].

Reflection Removal: Previous work on reflection removal has focused on decomposing the image into diffuse and specular components by using a polarization filter, changing focus, or chang-

ing viewpoint [Nayar et al. 1997; Schechner et al. 2000; Farid and Adelson 1999; Szeliski et al. 2000]. Levin et al. [2004] described a belief propagation based method to minimize the number of edges in the reflection-free decomposition of a single image. We propose a method to remove the reflections caused by a flash by using an ambient image.

Flash Settings on Cameras: On-board sensors allow modern cameras to automatically select the flash intensity and exposure settings based on aggregate measurements of scene brightness and distance. For example, Canon’s A-TTL (through-the-lens) method uses a pre-flash and a light sensor on the flash unit to compute the illumination needed for the scene. Canon’s E-TTL II technology [Canon] provides greater flash control by using only those metering areas that have a small difference between the ambient and the pre-flash readings. In the Nikon-3D system, as the camera focuses, the lens provides camera-to-subject distance which is then used to determine the best flash intensity. However, in all these cases the selected flash and exposure settings do not ensure that all regions of the scene (potentially having a large dynamic range) are adequately lit and captured. We show how current cameras may be programmed to adaptively and efficiently select these parameters to adequately capture a scene using a minimum number of images.

In Section 3, we present the flash imaging model, the image gradient coherence model and the gradient projection scheme. We use these concepts to remove flash artifacts in a variety of scenarios and present a technique to achieve uniform illumination in the presence of large depth variations in Section 4. In Section 5, we analyze the two-dimensional flash-exposure space and describe how the next best picture can be taken. This is followed by details on implementation in Section 6. We conclude the paper with a discussion on future directions.

3 Gradient Coherence and Projection

We first describe the flash imaging model and show how the flash and ambient images are related to scene reflectance, surface geometry and distance from the camera. We then discuss the gradient coherence model relating gradients in the flash and ambient images, followed by the gradient projection scheme.

3.1 Flash Imaging Model

The scene radiance captured in flash photography for a static scene is a linear combination of radiance due to the flash and the ambient illumination. Let Φ be the flash radiance map, or the image strictly due to a unit flash intensity. The flash-only image F is Φ scaled by the flash intensity P . Let α be the ambient radiance map, or the ambient image captured for unit exposure time. The ambient image A is α scaled by the exposure time E . The scale factors P and E are constant for all pixels. The irradiance map of a linear response camera is given by

$$I = F + A = \Phi P + \alpha E. \quad (1)$$

We assume that the flash duration, usually 1 millisecond, is significantly smaller than the exposure time E (usually 10-100 milliseconds). Note that since the camera and the flash are fixed in position, (1) is valid independent of scene reflectance (diffuse, specular or transparent objects), geometry (near or distant) and medium (air, fog, water or glass). Let us now examine the Φ and α terms in detail. The image irradiance, L , at a point with bidirectional reflectance distribution function, Ψ , in the direction ω_r , is given by

$$L(\omega_r) = \int_{\Omega} \Psi(\omega_i, \omega_r) L(\omega_i) \cos \theta_i d\omega_i, \quad (2)$$

where θ_i is the angle between the surface normal and the incident illumination direction. The incident irradiance term $L(\omega_i)$ includes

flash ($L^f(\omega_i)$) and ambient ($L^a(\omega_i)$) terms within the hemisphere, Ω . Specifically, for a diffuse object with reflectance ρ , the component due to the ambient illumination is

$$\alpha = \int_{\Omega} \left(\frac{\rho}{\pi} \right) L^a(\omega) \cos \theta d\omega = \frac{\rho B}{\pi}, \quad (3)$$

where $B = \int_{\Omega} L^a(\omega) \cos \theta d\omega$ is the aggregated ambience. Thus α depends on scene reflectivity, ρ , and the ambient illumination. The flash component is

$$\Phi = \frac{\rho \cos \theta_F}{\pi d^2}, \quad (4)$$

which depends on ρ , distance from the flash, d , and the angle between the flash direction and the surface orientation, θ_F . Thus, we expect points far from the camera or dark objects to have a low Φ value and regions near the camera and bright objects to have a high Φ value.

3.2 Gradient Orientation Coherence

We seek to capture properties of the image that remain invariant under change of lighting (**illumination invariants**) and use them to detect flash artifacts such as reflections and hot spots. We propose a coherence model based on the orientation of the image gradient vector. We use the observation that the orientation of image gradients remains stable under variable illumination if the gradients are due to local changes in reflectance and geometric shape and not due to local changes in illumination [Lichtenauer et al. 2004; Finlayson et al. 2004]. Chen et al. [2000] showed that for an object with Lambertian reflectance, discriminative functions that are invariant to illumination do not exist. However, the direction of the image gradients is insensitive to changes in the illumination to a large extent. Though we have only two illumination conditions, flash is a special kind of illumination given its proximity to the camera center. We assume that except at depth edges, the shadows induced by the flash are very few [Raskar et al. 2004a]. Hence, in general, we expect the gradient orientations in the flash image to be coherent with the gradient orientations in the ambient image, except at regions with flash artifacts and ambient shadow edges. Note that the coherence is observed not merely at the intensity edges, but essentially at all the pixels. We exclude pixels with gradient magnitude below 0.005 in the image with values $[0, 1]$, since the gradient directions at very low magnitude are unstable. This coherence may not hold if both illumination and reflectance gradients are small (e.g. on a diffuse tabletop). However, for smoothly varying illumination, presence of any reflectance gradient will boost the coherence. Let M denote the **gradient orientation coherency map** between Φ and α :

$$M = |\nabla \Phi \bullet \nabla \alpha| / (|\nabla \Phi| |\nabla \alpha|). \quad (5)$$

$M(x, y)$ encodes the angular similarity of flash and ambient image gradient vectors at pixel (x, y) . Thus, gradient orientation coherence indicates that although flash and ambient image intensities are quite different, in general, the gradient vectors in the flash and ambient images have the same orientation but different magnitudes. In other words, these gradients are related by an unknown scalar k :

$$\nabla \Phi = k \nabla \alpha. \quad (6)$$

The scalar k is different for each pixel and is independent of the reflectance. Hence, k will be relatively smooth over a smooth surface in the scene. k is positive at reflectance gradients but may become negative at depth edges and creases where the gradient polarity reverses.

Flash and Ambient Image Artifacts: In many cases, the flash image or the ambient image exhibits artifacts such as reflections, hot spots and specularities. These artifacts can be modeled as unknown

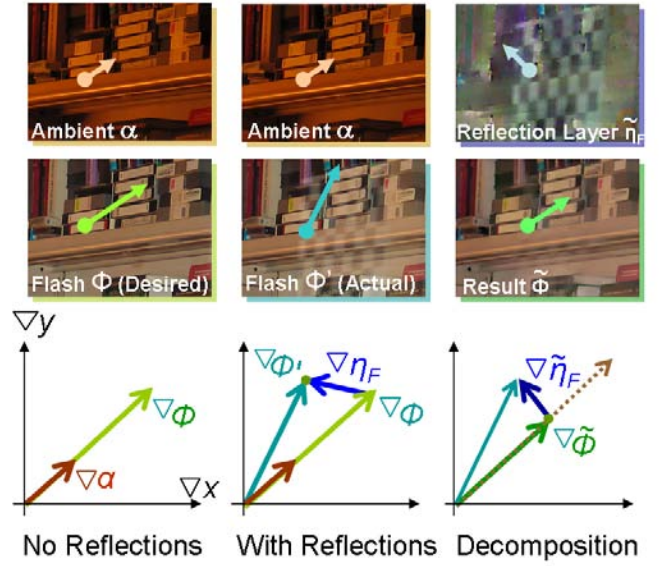


Figure 2: Overview of the reflection removal approach. The top two rows show images of an office scene behind a glass window. We took two flash images: with and without a printed checkerboard outside the office. The checkerboard is reflected on the glass window when a flash is used. (Left) The flash and ambient gradient vectors have a coherent orientation if there are no reflection artifacts. (Middle) Flash gradient $\nabla \Phi$ is corrupted by noise $\nabla \eta_F$ due to the reflected checkerboard. (Right) By removing the component of the noisy flash gradient vector $\nabla \Phi'$ perpendicular to the ambient gradient vector $\nabla \alpha$, the visual effect of noise can be significantly reduced. The computed reflection layer corresponding to the removed component $\nabla \tilde{\eta}_F$ is also shown.

noise to generate new gradients $\nabla \alpha'$ and $\nabla \Phi'$ as follows (see Figure 2):

$$\begin{aligned} \nabla \alpha' &= \nabla \alpha + \nabla \eta_A, \\ \nabla \Phi' &= \nabla \Phi + \nabla \eta_F = k \nabla \alpha + \nabla \eta_F. \end{aligned} \quad (7)$$

Decomposing the single flash image Φ' into two images, Φ and η_F , is an ill-posed problem. We have two equations and four unknowns ($k, \nabla \alpha, \nabla \eta_A$ and $\nabla \eta_F$). But it is possible to recover the undistorted components from this under-constrained problem by analyzing the gradient coherence and taking a vector projection *without* explicitly estimating k .

3.3 Gradient Projection

We show that by removing the component of the noisy gradient vector perpendicular to the signal gradient vector, the visual effect of noise can be significantly reduced. Let us first analyze the effect of rotating the gradients of an image by an angle ϕ . Let $G = [G_x, G_y]$ denote the gradient field of an image I . At each pixel, the gradients are rotated by ϕ to generate a new gradient field $G' = [G'_x, G'_y]$ given by

$$\begin{bmatrix} G'_x \\ G'_y \end{bmatrix} = \begin{bmatrix} \cos(\phi) & -\sin(\phi) \\ \sin(\phi) & \cos(\phi) \end{bmatrix} \begin{bmatrix} G_x \\ G_y \end{bmatrix}$$

Let I' denote the image reconstructed from $[G'_x, G'_y]$, which is obtained by solving a Poisson equation $\nabla^2 I' = \text{div}(G')$, where div

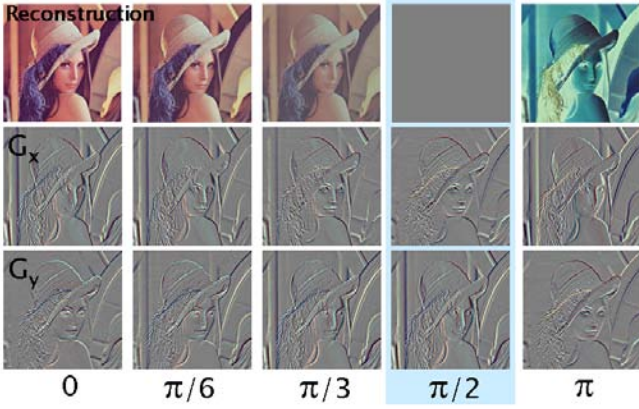


Figure 3: Poisson reconstruction after rotating the image gradient vectors. Notice that at $\varphi = \pi/2$, $[G'_x, G'_y] = [-G_y, G_x]$ and the image reconstructed after rotating each gradient vector by $\pi/2$ is zero.

denote the divergence. The divergence of G' is given by

$$\begin{aligned} \text{div}(G') &= \frac{\partial G'_x}{\partial x} + \frac{\partial G'_y}{\partial y} \\ &= \cos \varphi \left(\frac{\partial G_x}{\partial x} + \frac{\partial G_y}{\partial y} \right) + \sin \varphi \left(\frac{\partial G_x}{\partial y} - \frac{\partial G_y}{\partial x} \right), \quad (8) \\ &= \cos \varphi \text{div}(G) - \sin \varphi \text{curl}(I) = \cos \varphi \text{div}(G), \end{aligned}$$

since the curl of a scalar field is always zero. Thus, by rotating the image gradients by φ , the divergence of the gradient field decreases by a factor of $\cos \varphi$. Hence, at $\varphi = \pi/2$, the divergence is zero for any image. Figure 3 shows images reconstructed by rotating the gradients of the Lena image by different angles. Note that at $\varphi = \pi/2$, the Poisson equation $\nabla^2 I' = \text{div}(G')$ reduces to the Laplace equation $\nabla^2 I' = 0$. The reconstruction then depends on the boundary conditions. In our implementation, we zero-pad the images. Thus the reconstruction at $\varphi = \pi/2$ is zero everywhere, as shown in Figure 3. Since the reconstruction from gradient components orthogonal to the image gradients is zero, we can remove those components without effecting the visual content of the image. Let \rightarrow denote the projection operator. The gradient field obtained by the projection of the flash image gradients onto the ambient image gradients is given by

$$\nabla \tilde{\Phi} = (\nabla \Phi' \rightarrow \nabla \alpha) = \nabla \alpha (|\nabla \Phi' \bullet \nabla \alpha| / \|\nabla \alpha\|^2). \quad (9)$$

Note that the removed component $\nabla \tilde{\eta}_F = \nabla \Phi' - \nabla \tilde{\Phi}$ is orthogonal to $\nabla \Phi$ (see Figure 2). Thus we can remove the component of noise orthogonal to $\nabla \Phi$. However, we cannot eliminate the component of $\nabla \eta_F$ that is along $\nabla \Phi$. At the same time, as seen in the Lena images in Figure 3, small rotations of the gradients only reduces the effective magnitude of the image intensities. Therefore, minor variations in the undistorted $\nabla \alpha$ and $\nabla \Phi$ will simply reduce the effective magnitude in the reconstructed flash image $\tilde{\Phi}$ while preserving the visual information.

Gradient Direction Reversal at Depth Edges: At times, the angle between the flash and the ambient image gradient, λ , can be greater than $\pi/2$. This can happen at those depth edges where the flash and ambient image gradients have opposite orientation or when the magnitude of the noise gradient is large. In such cases, the direction of $\nabla \tilde{\Phi}$ needs to be reversed as shown in Figure 4. We assume that the high magnitude gradients of signal and noise do not overlap. We define a binary mask M' which is 0 at all pixels where both ambient and flash image gradients have high magnitude and $\lambda > \pi/2$. At those pixels, we reverse the direction of the projected gradient vector.

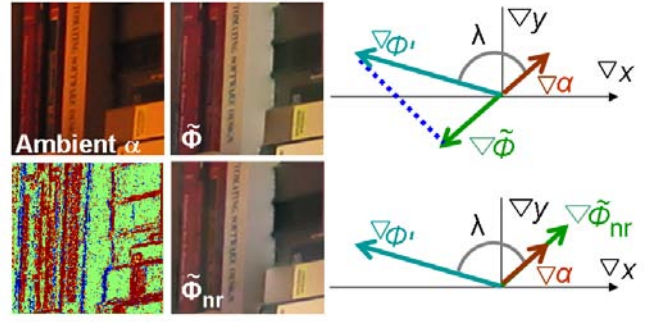


Figure 4: Gradient direction reversal at depth edges. (Left) The top-left patch of the ambient image shown in Figure 2, and the corresponding $\cos(\lambda)$ map (color-coded, red=1, blue=-1). Negative values mostly correspond to depth edges with opposite gradient orientation in flash and ambient images. (Middle) Reconstructed images where the direction of the projected gradient was reversed/not-reversed ($\tilde{\Phi}/\tilde{\Phi}_{nr}$) at depth edges with $\lambda > \pi/2$. Notice the jagged depth edges in $\tilde{\Phi}_{nr}$. (Right) Flash and ambient image gradients for $\lambda > \pi/2$. At depth edges, the direction of the projected gradient is reversed to obtain the correct $\nabla \tilde{\Phi}$ (top). If the direction is not reversed (bottom), depth edges are smeared in the reconstructed image.

4 Flash Image Artifacts

We now show how the gradient coherency map and the gradient projection scheme can be used to remove the flash artifacts. In general, the gradient coherency map M is used as a guide to combine the flash and ambient image gradients and the gradient projection scheme is used to remove unwanted reflections of objects lit by the flash.

4.1 Reflection Removal

We describe the gradient projection-based removal of flash artifacts in a variety of scenarios and also demonstrate some of the limitations of the proposed approach. In photographing through transparent layers such as coatings on paintings, glass separators or windows, flash can create undesirable reflections of objects or a hot spot. In some cases the flash is intended to illuminate what is behind the glass (e.g. fish in a tank) but creates reflections of objects in front of the glass. In some other cases, the goal is to illuminate objects in front of the glass, such as a person in front of a painting, but the flash creates a hot spot. An ambient image can be used in these cases to reduce the artifacts. In some cases even the ambient image has reflections; in certain conditions these can be reduced by using a flash image.

Showroom Window: In this scenario, the flash is intended to illuminate what is behind the glass but also creates reflections of objects in front of the glass (Figure 5). Hence, $\eta_A = 0$, η_F is non-zero and $\nabla \Phi$ and $\nabla \alpha$ are coherent at all pixels except where the flash reflections are present. The goal is to create a reflection-free flash image. To achieve this, we obtain a new gradient field $\nabla \Phi^*$ by projecting the flash image gradients onto the ambient image gradients.

$$\nabla \Phi^* = (\nabla \Phi' \rightarrow \nabla \alpha). \quad (10)$$

The artifact-free flash image is obtained by integrating $\nabla \Phi^*$ (Figure 5). This example also shows some of the limitations of this approach. Note that the reflections are still visible in the final result on the homogeneous white board due to the lack of reliable ambient image gradients there. One might argue that the ambient image is



Figure 5: Showroom Window. Flash illuminates objects behind the glass but also creates undesirable reflections. Reflections are removed by taking the projection of the flash image gradients onto the ambient image gradients. (Left) Ambient image. (Middle) Flash image. (Right) Enhanced flash image with reflections removed.

the best of the three images, but in the ambient image, the hair disappears and the red brick texture has low contrast. The flash image enhances the contrast.

Museum: Suppose we want to take a photograph of a person in front of a painting in a museum or an art gallery. The painting is well-lit by the ambient illumination, but a flash is required to illuminate the dimly lit person. The flash, however, creates a hot spot as shown in Figure 1 (left sub-figure). Thus, $\eta_A = 0$, η_F is non-zero at the hot spot and $\nabla\Phi$ and $\nabla\alpha$ are coherent everywhere except at the hot spot. We detect the hot spot using a saturation map. We first normalize the image intensities to lie between $[0, 1]$. For each pixel (x, y) , the saturation weight w_s corresponding to the normalized intensity I is given by

$$w_s(x, y) = \tanh(\sigma * (I(x, y) - \tau_s)), \quad (11)$$

where $\sigma = 40$ and $\tau_s = 0.9$. The saturation map is then normalized to lie between $[0, 1]$. Note that a threshold based scheme will not be effective in this case as the hot spot spreads over an area and is not a point or sharp line specularity. We obtain the new gradient field $\nabla\Phi^*$ by linearly weighting the flash and ambient image gradients using the saturation map and the coherency map as

$$\nabla\Phi^* = w_s \nabla\alpha + (1 - w_s)(M \nabla\Phi' + (1 - M) \nabla\alpha). \quad (12)$$

The artifact-free flash image is obtained by integrating $\nabla\Phi^*$. Note that a professional photographer with a pointable flash would use bounce-flash off the ceiling to avoid the hot spot. However, our technique works with a consumer level camera.

Observation Deck: Consider the common scenario of taking a picture of a person from inside a glass enclosed room at night (Figure 6). A 'night scene' mode for the camera is commonly used in such scenarios. The person is not well-lit in the ambient image, and although the flash illuminates the person, it creates reflections of the objects inside the room. Clearly, in such scenes the flash does not effect distant buildings. Thus, one can add the flash and the ambient image to get an image H , in which the person is well-lit. However, the reflections of the room objects are still present in H . To remove the reflections in H , we take the projection of ∇H onto $\nabla\alpha$. However, the projection will be unreliable at the pixels where the ambient image is underexposed. We create an underexposed map where for each pixel (x, y) , the weight w_{ue} is defined as

$$w_{ue}(x, y) = 1 - \tanh(\sigma * (I(x, y) - \tau_{ue})), \quad (13)$$

where I is the normalized intensity. The underexposed map is normalized to lie between $[0, 1]$. We use $\sigma = 40$ as before and $\tau_{ue} = 0.1$. The new gradient field $\nabla\Phi^*$ is obtained as

$$\nabla\Phi^* = w_{ue} \nabla H + (1 - w_{ue})(\nabla H \rightarrow \nabla\alpha). \quad (14)$$



Figure 6: Observation Deck. Flash is used to illuminate the objects in front of the glass but causes reflections. Objects behind the glass are not affected by the flash. (Left) Ambient image. (Middle) Flash image. (Right) Enhanced flash image with reflections removed.

The enhanced image is obtained by integrating $\nabla\Phi^*$. Note that if the reflections fall on a region which is dark in the ambient image, it cannot be removed due to the lack of information. Since the flash-lit parts can be segmented in this scenario, the flash and ambient images can also be fused using the method in [Perez et al. 2003]. However, their method requires the use of an explicit image mask whereas ours is automatic.

Self Reflection: We show that a flash image can also be used to remove reflections in an ambient image. For example, while photographing a well-lit painting, the photographer is often reflected in the painting as seen in Figure 1 (right sub-figure). To avoid reflections, one could use a low-exposure flash but it creates a hot spot. The goal here is to recover the reflection-free ambient image without the hot spot. To remove ambient reflections, we take the projection of the ambient image gradients onto the flash image gradients. Since the projection onto the flash image gradients cannot be taken at the hot spot where the flash image gradients are unreliable, we assume no reflections and use the original ambient image gradients at those pixels. The new gradient field is obtained as

$$\nabla\alpha^* = w_s \nabla\alpha' + (1 - w_s)(\nabla\alpha' \rightarrow \nabla\Phi'). \quad (15)$$

Integrating this field gives us the artifact-free ambient image. The *noise* signal or the reflection layer can be obtained by taking the orthogonal component of the projection. Thus the gradient field of the reflection layer is obtained as

$$\nabla\tilde{\eta}_A = (1 - w_s)(\nabla\alpha' - (\nabla\alpha' \rightarrow \nabla\Phi')). \quad (16)$$

Figure 1 also shows the computed reflection layer. Note that the photographer is clearly visible in the reflection layer despite the colors of his shirt and the background being similar. Also, notice the folds of the photographer's shirt at the bottom and the texture of the hands in the recovered reflection layer. However, co-located artifacts such as ambient reflections at the location of the flash hot spot cannot be removed using this approach.

4.2 Depth Compensation

Since the flash intensity falls off with the distance from the camera, flashes produce a tunnel effect where brightness decreases quickly with depth. Thus, distant objects are poorly lit by the flash as compared to the nearby objects. We show that the ratio of the flash and ambient images provides information about the depth and orientation of surfaces in the scene and can be used to scale the image intensities of distant objects in the flash image. Let β denote the ratio of the flash and the ambient radiance map. Using (3) and (4), one obtains $\beta = \Phi/\alpha = \cos\theta_F/(d^2B)$. Note that β is independent of the reflectance and can be regarded as a *depth-orientation map*. For distant objects, β will be small, and for nearby objects β will be large. If we assume that the ambient illumination B is uniform

or low-frequency, we can enhance the flash image to compensate for the $d^2/(\cos \theta_F)$ attenuation using β . To achieve depth compensation, we propose to scale the flash image gradients by $1/\beta$, thus attenuating gradients for objects near the camera and accentuating gradients for objects far away from the camera. However, the flash image gradients should be scaled only when they are coherent with the ambient image gradients to avoid artifacts due to ambient illumination. Thus, the new gradient field $\nabla\Phi^*$ is obtained as

$$\nabla\Phi^* = M\nabla\Phi/\beta + (1-M)\nabla\Phi.$$

Figure 7 shows the results. Other intensity normalization techniques such as gamma correction, contrast reduction or nonlinear scaling of gradients [Fattal et al. 2002] (shown in the figure) changes the overall appearance of the images. In contrast, by exploiting the β map, our technique applies local scaling to the flash image.

5 Flash-Exposure Space

Current cameras use onboard sensors and processing to coarsely estimate the flash level and the exposure settings. But these estimates are based on aggregate measurements and lead to the common problem of over-illumination or under-illumination. It is difficult to find a single flash intensity value that can light up distant or dark objects without saturating nearby or bright objects. Also, the quality of the flash/no-flash images may be limited in terms of dynamic range. Figure 8 shows an example of such a HDR scene. In such cases, we advocate using several images taken under different flash intensities and exposures to build a *flash-exposure HDR* image.

5.1 Flash-Exposure HDR Image

Suppose we capture $N_F * N_A$ images with N_F different flash intensities and N_A different exposures. After linearizing the camera response, the ambient radiance map, α , can be estimated using traditional HDR methods by using the N_A samples along the E axis which have different exposures [Mann and Picard 1995; Debevec and Malik 1997]. The flash radiance map, Φ , can be estimated using a *flash-HDR* method by using the N_F samples parallel to the P axis which have different flash intensities at a fixed exposure E^* . However, the above methods either use the 1-D exposure space or the 1-D flash space to estimate the Φ and α radiance maps. Instead, we propose to construct a flash-exposure HDR image by using samples from the 2-D flash-exposure space as follows.

For each pixel (x, y) in all $N_F * N_A$ images, we can rewrite (1) as

$$I(x, y)^{i,j} = \Phi(x, y)P^i + \alpha(x, y)E^j, \quad (17)$$



Figure 7: Dinner Table. Non-uniform illumination in a scene with large depth variation. Our method selectively scales the flash image gradients such that the brightness falloff with the distance from the camera is minimized. A direct tone-mapping procedure leads to loss in contrast.

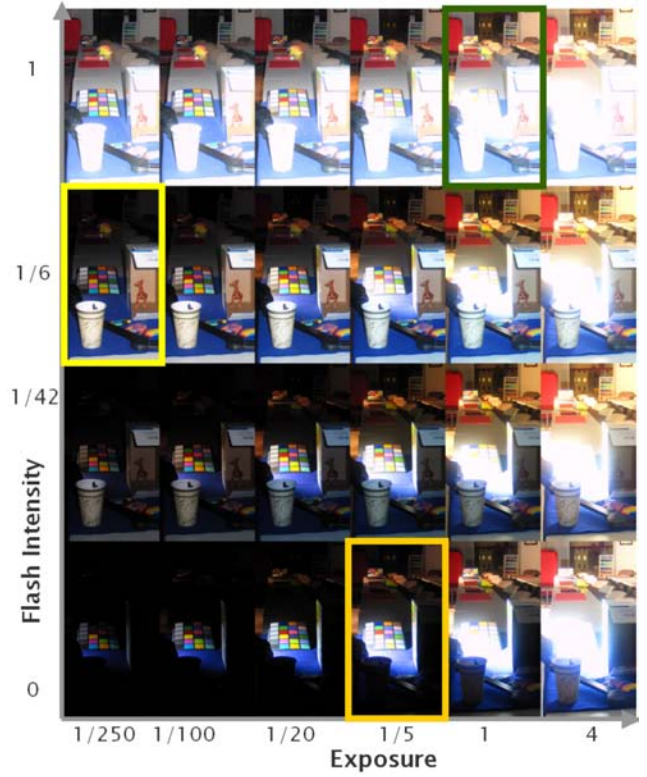


Figure 8: Two-dimensional space of flash intensity and exposure parameters. In a scene with large variations in depth, illumination and reflectance, one may need to take multiple pictures to estimate the (Φ, α) parameters at each pixel. Instead of taking samples along the E or the P axis, our adaptive sampling method minimizes the number of required samples. In this case, we start with the images shown within the yellow and the orange boxes. The image corresponding to the next best (P, E) settings is shown within the green box.



Figure 9: Flash-Exposure HDR map. (Left) Ambient HDR map α . (Middle) Flash HDR map Φ . (Right) Combined flash-exposure HDR map obtained by fusing the α and Φ maps has more information than the individual Φ and α maps. Notice the barely visible zebra inside the box in the α map and the unlit windows at the back in the Φ map. All such parts of the scene are captured well in the combined flash-exposure HDR map.

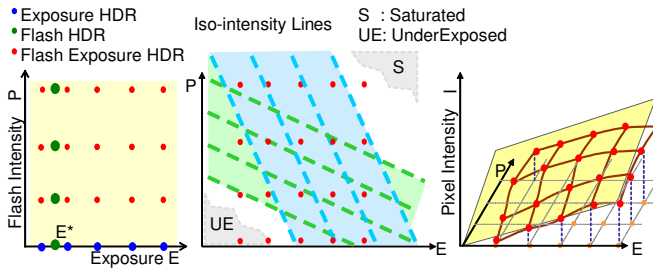


Figure 10: Flash-Exposure Space. (Left) Samples corresponding to exposure HDR, flash HDR and flash-exposure HDR. (Right) Parallel iso-intensity contours for two pixels. These pixels will be saturated in the images corresponding to the P - E values in the region S and will be underexposed in the images corresponding to the P - E values in the region UE . (Right) The estimated Φ and α values for a pixel correspond to the slopes of the best-fit plane in the P - E space.

where $i = 1 \dots N_F$ and $j = 1 \dots N_A$. Thus, for each pixel we solve a system of $N_F * N_A$ linear equations to obtain a least square estimate of the two unknowns $\Phi(x,y)$ and $\alpha(x,y)$. A better weighted least square estimate can be obtained by weighting each equation by $\Gamma(I(x,y)^{i,j})$, where $\Gamma(I(x,y)^{i,j})$ is the SNR for the intensity $I(x,y)^{i,j}$ at the current ISO setting [Kodak 2001]. Figure 9 shows our estimates of the Φ and α maps for the images shown in Figure 8. We intentionally use a simple logarithmic mapping to display the HDR maps in Figure 9 (thus they may look washed out). A combined flash-exposure HDR map can be obtained by fusing the Φ and α radiance maps. We use a method similar to the one proposed in [Raskar et al. 2004b] by taking the maximum of the gradients from both, thus preserving maximum contrast. However, our technique has one important difference with respect to the above previous work. We account for the gradient direction reversal at the depth edges from the Φ map to the α map using the mask M' (see section 3.3) while choosing the maximum gradient. At all pixels where $M' = 0$, we keep the original flash image gradients. Let T denote a mask such that T is 1 if $|\nabla\Phi| \geq |\nabla\alpha|$ and 0 otherwise. The new gradient field ∇R is obtained as

$$\nabla R = M' (T \nabla\Phi + (1-T) \nabla\alpha) + (1-M') \nabla\Phi.$$

The combined flash-exposure HDR image (Figure 9) is obtained by integrating ∇R .

5.2 Flash-Exposure Walk

Capturing multiple images for HDR imaging may not be practical due to the static scene/camera assumption. Can we minimize the number of images required to capture the entire dynamic range of the scene? Or, given a budget of N images, what are the best exposure and flash settings for those N images so that a maximum region of the scene is captured with good SNR? Our goal is to design a *scene-adaptive* technique. Grossberg & Nayar [2003] have looked at taking non-uniform samples along the exposure axis to account for a camera's non-linear radiometric response. Their camera-response-adaptive technique is complementary to our scene-adaptive technique.

We analyze the two-dimensional flash intensity-exposure space (P - E space) and propose a method for adaptively sampling this space to estimate the best flash-exposure settings. By understanding the nature of the P - E space, one can intelligently walk in this space. The P - E space has a wealth of information (Figure 10). The iso-intensity contour for an intensity value μ at a pixel (x,y) corresponds to the line in the P - E space given by



Figure 11: Optimal flash intensity and exposure (P,E) settings for the next picture. (Left) Pixels shown in white are captured with high SNR in the first two images (shown within the yellow and the orange boxes in Figure 8). (Middle) Pixels shown in white are captured with high SNR in the third image (shown within the green box in Figure 8). (Right) Pixels (in white) which are captured with high SNR using all three images. Note that almost all the pixels which are not captured with good SNR in the first two images (windows at the back, part of the table between the first and the second Macbeth chart) are captured with good SNR in the third image.

$\Phi(x,y)P + \alpha(x,y)E = \mu$. These lines will be vertical for distant objects not illuminated by the flash, horizontal for nearby objects not lit in the ambient image, and in general slanted for objects lit in both flash and ambient images. For diffuse reflection, the slope is approximately equal to $\beta = \cos \theta_F / (d^2 B)$, which is independent of the reflectance.

Minimum Sample Walk in the P - E Space: We start with two pictures, a no-flash image and a low-exposure flash image, both captured using the exposure and flash intensity parameters suggested by the camera's on-board sensor (such settings are common in the 'night scene mode'). From these two pictures, we obtain low quality estimates of the Φ and α values for each pixel. To find the next best flash and exposure settings, we first predict the intensity I of each pixel using the current estimates of Φ and α for a candidate (P,E) using (1). Then, we compute the SNR for each pixel. Since our goal is to sense both the flash and the ambient components at high SNR, the SNR for a pixel is the minimum of the SNR of the component intensities, i.e., $\min(\Gamma(\Phi P), \Gamma(\alpha E), \Gamma(I))$. We select the (P,E) values that maximize the SNR over all the pixels. However, we exclude those pixels which have already been sampled with good SNR in the previous pictures. Otherwise, well-exposed pixels will always dominate the estimation of the (P,E) parameters, keeping the estimates close to the current (P,E) values. Black objects in the scene may effect the estimation of the flash-exposure settings. We ignore black objects by ignoring very low intensity pixels at high flash intensity and exposure settings in the above analysis. With each new measurement, we update the estimates of Φ and α . The above procedure is continued until N pictures have been taken or all the pixels have been measured with good SNR.

We simulated the adaptive sampling scheme for the images shown in Figure 8. The images corresponding to the first two samples are shown within the yellow and the orange boxes. Our adaptive sampling technique suggests the next best picture to be the one shown within the green box. Figure 11 shows the pixels that are captured with high SNR in the first two pictures. The third picture captures most of the remaining pixels with good SNR.

6 Implementation

A Canon G3 camera, which supports 3 levels of flash intensities, was used for capturing the images. The images were captured with ISO setting varying between 50 and 200. The exposure values for the ambient images shown in Section 4 were either 1/30 or 1/10 second. For the flash images, the exposure time varied between 2-4 milliseconds. We assume no mis-registration between the flash and ambient images. The relative flash intensities were computed by weighted averaging of intensities observed using a Macbeth chart. Basic edge-preserving smoothing was performed on the input images to reduce image acquisition noise. A series designed filter of radius 5 was used for computing the image gradients [Jahne 1993]. Reconstruction of image intensities from the gradients amounts to solving a Poisson equation [Fattal et al. 2002]. For solving the Poisson equation, a sine transform based method [Press et al. 1992] was used. The images were zero-padded on both sides, and Dirichlet boundary conditions instead of Neumann boundary conditions were used to avoid the scale/shift ambiguity in the gradient reconstruction.

All three color channels were treated separately. However, color coherency was not assumed between the different channels. To maintain color consistency across the channels, the ambient image was white-balanced using the approach described in [Petschnigg et al. 2004]. Better results were obtained when using the YUV color space as compared to the RGB color space. Our technique does not require any human intervention. Run-time for a 1 Mega-pixel image is about 6 seconds using our unoptimized Matlab implementation.

7 Discussion

There are several possible extensions to our work. One could modify the ‘night scene’ mode available on point and shoot cameras to give two pictures instead of one. The exposure time for capturing a flash image needs to be only a few milliseconds, and hence flash images (each with a different flash intensity) can be easily captured in quick succession. This can be treated as flash intensity auto bracketing or flash-exposure auto bracketing. Flashes cause backscatter in fog and water, resulting in severe image degradation. However, the flash imaging model remains valid and our technique can be used to improve the flash images. Taken as a whole, a study of flash-exposure-aperture space can lead to further research in computational approaches to overcome problems in focus and motion blur.

8 Conclusion

We have presented techniques to improve flash photography and have addressed three well-known problems: over-illumination or under-illumination at a given flash intensity, reflections or highlights, and attenuation over depth. We reduce these artifacts by exploiting information in the ambient image. To remove reflections and highlights, one might think that higher-level prior knowledge or global image features are required. But we have presented a simple technique based on the local gradient analysis. Nevertheless, a global scheme may improve the results. To overcome the limited dynamic range in the flash and ambient images, we propose adaptively sampling the flash-exposure space. By understanding the flash-exposure space, one can design better cameras and develop novel re-synthesis effects. Current cameras already use on-board sensors and processing to coarsely estimate the flash level and exposure settings. We hope our methods can further improve pre-capture parameter selection and post-capture processing of the images.

Acknowledgements We would like to thank the anonymous reviewers and several members of the Mitsubishi Electric Research

Labs for their suggestions. We also thank Rama Chellappa, Joe Marks and Karhan Tan for helpful discussions and support.

References

- AGARWALA, A., DONTCHEVA, M., AGRAWALA, M., DRUCKER, S., COLBURN, A., CURLESS, B., SALESIN, D., AND COHEN, M. 2004. Interactive digital photomontage. *ACM Transactions on Graphics* 23, 3 (Aug.), 294–302.
- CANON. [Http://www.cps.canon-europe.com/kb/](http://www.cps.canon-europe.com/kb/).
- CHEN, H., BELHUMEUR, P., AND JACOBS, D. 2000. In search of illumination invariants. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, vol. 1, 254–261.
- DEBEVEC, P. E., AND MALIK, J. 1997. Recovering high dynamic range radiance maps from photographs. In *Proc. of the 24th annual conference on Computer graphics and interactive techniques*, 369–378.
- EISEMANN, E., AND DURAND, F. 2004. Flash photography enhancement via intrinsic relighting. *ACM Transactions on Graphics* 23, 3 (Aug.), 673–678.
- FARID, H., AND ADELSON, E. H. 1999. Separating reflections and lighting using independent components analysis. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, vol. 1, 1262–1267.
- FATTAL, R., LISCHINSKI, D., AND WERMAN, M. 2002. Gradient domain high dynamic range compression. *ACM Transactions on Graphics* 21, 3, 249–256.
- FINLAYSON, G., HORDLEY, S., AND DREW, M. 2002. Removing shadows from images. In *Proc. of European Conf. on Computer Vision*, vol. 4, 823–836.
- FINLAYSON, G. D., DREW, M. S., AND LU, C. 2004. Intrinsic images by entropy minimization. In *Proc. of European Conf. on Computer Vision*, vol. 3, 582–595.
- GROSSBERG, M. D., AND NAYAR, S. K. 2003. High dynamic range from multiple images: which exposures to combine? In *Proc. of ICCV Workshop on Color and Photometric Methods in Computer Vision*.
- JAHNE, B. 1993. *Spatio-temporal image processing, theory and scientific applications*, vol. 751 of *Lecture Notes in Computer Vision*. Springer-Verlag.
- KODAK, 2001. CCD image sensor noise sources. Application note MPT/PS-0233.
- LEVIN, A., ZOMET, A., AND WEISS, Y. 2004. Separating reflections from a single image using local features. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, vol. 1, 306–313.
- LICHTENAUER, J., REINDERS, M., AND HENDRIKS, E. 2004. Influence of the observation likelihood function on particle filtering performance in tracking applications. In *Sixth IEEE Int'l. Conf. on Automatic Face and Gesture Recognition*, 767–772.
- MANN, S., AND PICARD, R. W. 1995. Being undigital with digital cameras: extending dynamic range by combining differently exposed pictures. In *Proc. of IS&T 48th annual conference*, 422–428.
- NAYAR, S. K., AND MITSUNAGA, T. 2000. High dynamic range imaging: spatially varying pixel exposures. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, vol. 1, 472–479.
- NAYAR, S. K., FANG, X.-S., AND BOULT, T. 1997. Separation of reflection components using color and polarization. *Int'l. Journal of Computer Vision* 21, 3 (Feb.), 163–186.
- PEREZ, P., GANGNET, M., AND BLAKE, A. 2003. Poisson image editing. *ACM Transactions on Graphics* 22, 3, 313–318.
- PETSCHNIG, G., AGRAWALA, M., HOPPE, H., SZELISKI, R., COHEN, M., AND TOYAMA, K. 2004. Digital photography with flash and no-flash image pairs. *ACM Transactions on Graphics* 23, 3 (Aug.), 664–672.
- PRESS, W. H., TEUKOLSKY, S., VETTERLING, W. T., AND FLANNERY, B. P. 1992. *Numerical recipes in C: the art of scientific computing*. Pearson Education.
- RASKAR, R., TAN, K., FERIS, R., YU, J., AND TURK, M. 2004. Non-photorealistic camera: depth edge detection and stylized rendering using multi-flash imaging. *ACM Transactions on Graphics* 23, 3, 679–688.
- RASKAR, R., ILIE, A., AND YU, J. 2004. Image fusion for context enhancement and video surrealism. In *Proc. of NPAR*, 85–95.
- SCHECHNER, Y. Y., KIRYATI, N., AND BASRI, R. 2000. Separation of transparent layers using focus. *Int'l. Journal of Computer Vision* 39, 1 (Aug.), 25–39.
- SUN, J., JIA, J., TANG, C.-K., AND SHUM, H.-Y. 2004. Poisson matting. *ACM Transactions on Graphics* 23, 3, 315–321.
- SZELISKI, R., AVIDAN, S., AND ANANDAN, P. 2000. Layer extraction from multiple images containing reflections and transparency. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, vol. 1, 246–243.