Spatially Adaptive Photographic Flash

Rolf Adelsberger[1], Remo Ziegler[1], Marc Levoy[2], Markus Gross[1]

[1]ETH Zurich, Switzerland
[2]Stanford University, USA

Abstract

Using photographic flash for candid shots often results in an unevenly lit scene, in which objects in the back appear dark. We describe a spatially adaptive photographic flash system, in which the intensity of illumination varies depending on the depth and reflectivity of features in the scene. We adapt to changes in depth using a single-shot method, and to changes in reflectivity using a multi-shot method. The single-shot method requires only a depth image, whereas the multi-shot method requires at least one color image in addition to the depth data. To reduce noise in our depth images, we present a novel filter that takes into account the amplitude-dependent noise distribution of observed depth values. To demonstrate our ideas, we have built a prototype consisting of a depth camera, a flash light, an LCD and a lens. By attenuating the flash using the LCD, a variety of illumination effects can be achieved.

Categories and Subject Descriptors (according to ACM CCS):
I.3.2 [Computer Graphics]: Graphics Systems
I.3.3 [Computer Graphics]: Picture/Image Generation

1. Introduction

Digital cameras have become inexpensive, robust, and small. As a result, many users regularly carry one and expect it to take good pictures under a variety of shooting conditions. To accommodate dark scenes, most digital cameras include electronic flash - either integrated with the camera or in a detachable unit. More expensive units offer control over brightness, duration, and angular spread. Regardless of how these parameters are chosen, flash illumination always falls off sharply with distance from the camera, making most flash pictures look unnatural. More advanced flash units adapt the brightness to the front-most object, however, as a consequence more distant scene parts are often under-saturated. It is tempting to address this problem by capturing a flash-no-flash image pair, estimating depth as the ratio of the flash and no-flash images, then inversely scaling pixel intensities by this depth. However, separately scaling the intensities of individual pixels does not account for possible inter-reflections among the corresponding features in the scene. This can lead to physically invalid scenes, and hence to images that appear unnatural.

We solve this problem by changing the physical illumination of the scene on a pixel-by-pixel basis. Our system employs an infrared time-of-flight (ToF) rangefinder to determine the distance to each object and a modified video projector to modulate the illumination. While this approach is limited by the power of our light source, the illumination it produces is physically plausible, as are the inter-reflections it triggers in the scene, so our images look natural.

Using such a system it is possible to simulate a variety of virtual light sources. Unless the light source is a dramatic player in the scene, most photographers prefer sources that are diffuse, distant from the primary subject, and angled with respect to the direction of view. For interior shots, examples are ceiling lights or the sun passing through a side window. Since our spatially adaptive flash unit sits atop the camera, we cannot simulate lights that arrive from the side or above the subject. However, by modulating the illumination of each object as a function of its distance from the camera, we can simulate a light source located some distance behind the photographer. Illumination from such a source falls off quadratically with distance, but at a slower rate than from a flash unit mounted on the camera. In order to avoid unnatural shadows above the subject, we mount our physical light source above the camera, as for conventional flash.

(a) Regular flash: dark background.

(b) Computationally increased brightness: foreground becomes saturated, background suffers from noise.

(c) Spatially adapted flash (our results).

(d) Detailed view of Fig. 1(a): too dark background.

(e) Detail of Fig. 1(b): high noise level.

(f) Detail of Fig. 1(c): improved SNR.

(g) Mask used for Fig. 1(c)

Figure 1: *Applying SAFU to a natural scene (top row). Details on SNR (bottom row)*

Our method requires only one snapshot by the main camera, preceded by a single capture using the rangefinder. For scenes in which flash provides the bulk of the illumination, a second problem is that highly reflective objects may saturate the sensor, especially if they lie close to the camera. By capturing an additional pilot image to estimate object reflectance, we can control the spatial light modulator (in our prototype-implementation: a video projector) to reduce the illumination of these objects during the final image capture. One can treat this additional modulation as a sort of "physical tone mapping" - reducing the dynamic range of the scene by adjusting its illumination locally. In this paper we describe the hardware and algorithms of our system, and we demonstrate its use in situations where flash typically performs poorly, such as candid shots of human faces.

2. Related Work

Graphics and vision researchers have long recognized the importance of controlling illumination in computational photography methods. We can classify these methods according to the kind of illumination they employ:

Single Spatially Invariant Flash Petschnigg et al. [PSA*04] and Eisemann et al. [ED04] take two images - one with flash and one without, then use the low noise of the flash image to reduce noise in the no-flash image. For this purpose they employ a joint bilateral filter (a.k.a. cross-bilateral filter). However, using a spatially invariant flash on a scene with a large depth extent can lead to underexposed regions in the flash image. These dark regions are noisy, thereby diminishing the quality of the detail that can be transferred to the corresponding regions of the no-flash image. Using our spatially varying flash, we can ameliorate this problem.

Single Spatially Invariant Flash using Multiple Shots Several problems of flash images, such as strong highlights due to reflection of the flash by glossy surfaces, saturation of nearby objects, and poor illumination of distant objects are addressed by Agarwal et. al [ARNL05]. They capture a flash-no-flash image pair, estimate depth as the ratio of the flash and no-flash images, then inversely scale the gradients of each pixel in the flash image by this depth. The final image is computed by integrating this gradient field. However, separately scaling the intensities of individual pixels does not account for possible inter-reflections among features in the scene, as mentioned earlier. By contrast, our spatially adaptive flash applies different illumination along each ray. This produces a physically valid scene, and it requires capturing only one image rather than a pair of images.

Programmable Illumination Using a projector-like light source is a powerful alternative to spatially invariant flash illumination. Nayar et. al [NKGR06] show how a programmable flash can be used to separate an image into direct

and indirect reflections. However, they cannot compute how these reflections would change if only a part of the scene were lit - a necessary capability if one wishes to reduce illumination on nearby objects. Mohan et al. [MBW*07] presents an inexpensive method for acquiring high-quality photographic lighting of desktop-sized static objects such as museum artifacts. They take multiple pictures of the scene, while a moving-head spotlight is scanning the inside of a foam-box enclosing the scene. This requires a controlled environment, which is typically not possible for candid shots. Other systems for programmatically controlled illumination include time-multiplexed lighting [SNB07] and Paul Debevec's Light Stages [DHT*00]. However, these methods are not suitable for candid photography.

Dynamic Range Many methods have been proposed for improving the dynamic range of an imaging system. These systems typically compromise either temporal or spatial resolution, or they employ multiple detectors. Nayar et. al [NB03] provides a good summary of previous work. They also present a way to enlarge the dynamic range of a system by adding a spatial light modulator (LCD) or DMD [NBB06] to the optical path. This enables them to extend the dynamic range of the sensor by varying the exposure on a per-pixel basis. Our work is complementary to this method. It permits us to attenuate light that is too bright, while also providing a way to add light to regions that would otherwise be too dark.

3. System Overview

Our proposal can be partitioned into two variants: single-shot and multi-shot. In the former variant, a depth image can be acquired during the focusing of the color camera, thus we consider this a single-shot method. Fig. 1 was captured using this method. In the multi-shot variant, the first color image serves as a pilot image, from which we estimate the brightness of each feature in the scene. We use this information to further refine our spatially varying illumination during a final color image acquisition. We can also use this information to refine our depth image. Fig.12(c) was captured using this method.

Fig.2 shows an overview of our system. The single-shot and multi-shot approaches can be split into four steps, acquisition, depth filtering, reprojection, and lighting. We apply a median filter to remove outliers in depths returned by the ToF-camera, and a novel trilateral filter to reduce noise. The filtered depth data is then backprojected, triangulated, and rendered into the projector view using graphics hardware. This step performs warping and resolves visibility in real-time. To cope with inaccuracies in calibration or depth values, we apply a conservative depth discontinuity diffusion filter. This filter gives us depth values that are correct or slightly closer than the actual depth, thereby diminishing the visibility of possible misalignments. Next an attenuation mask is computed. This computation takes as input the filtered depths, the virtual light position, and a user defined artistic filter. This attenuation mask is loaded onto the LCD of the Spatially Adaptive Flash Unit (SAFU), and a final image is captured by triggering the flash and camera to fire.

In the multi-shot approach, depth filtering is carried out in two phases rather than one. In the first phase the depth filtering of the single-shot approach is applied to the depth data, after which it is warped into the color camera's view. In the second phase, we capture a color image, then apply a slightly modified joint bilateral filter [PSA*04] to refine the depth edges according to the color data. The resulting high resolution depth information is then reprojected into the projector view. Finally, the attenuation mask is computed as in the single-shot method, but based on a high resolution mesh instead of low-resolution. Additionally, reflectance information from a flash-no-flash pair of color images is used to refine the attenuation mask to reduce illumination of highly reflective objects.

The hardware setup and the required calibration are described in the remainder of this section. In the two following sections we describe, firstly, the entire system using a single-shot approach (cf. Sect. 4), and secondly, the additional benefits and applications using the multi-shot approach (cf. Sect. 5).

3.1. Hardware Setup

The setup consists of a Spatially Adaptive Flash Unit (SAFU) as well as a color camera. In our standard setup we position the SAFU, such that the location roughly corresponds to a conventional flash unit mounted on the hot shoe of the color camera (cf. Fig. 3). This allocation of the different components has been chosen to get as close as possible to a candid shot setting. However, it might be possible to trigger the flash remotely, but this would require a re-calibration each time the SAFU is moved relatively to the cameras.

3.1.1. SAFU

The SAFU consists of a lamp, a LCD panel, a lens and a depth camera based on time-of-flight technology. The conceptual setup is shown in Fig. 3a. In our current setup shown in Fig. 3b, we removed the light bulb from a projector (SANYO PLC-XW50) and replaced it by a Canon Speedlite 580EX II. In order to increase the intensity from the SAFU, we built a mirrored tube. This maximizes the light projected into the light path. The rest of the light path, including three LCD's, a lens, and the depth camera, are not modified. The alignment of the optical axis of the depth camera and the optical axis of the lens of the light are chosen to be as close as possible, in order to minimize the disparity between their views of the seen.

Time of Flight Camera

We employ a Swissranger ToF-camera from Mesa-Imaging with a resolution of 176×144 pixels, providing the depth

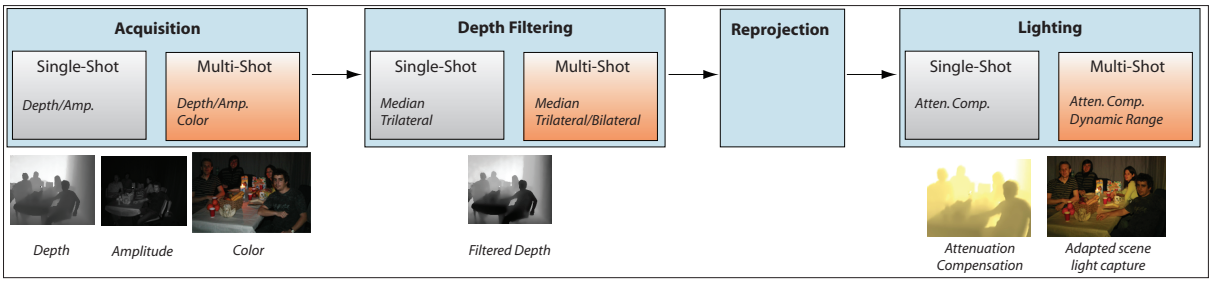


Figure 2: *The system can be split into two approaches, namely a single-shot and a multi-shot method. They consist of an acquisition, depth filtering, reprojection and lighting step.*

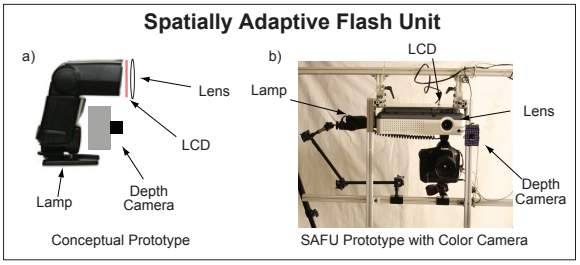


Figure 3: *a) The proposed SAFU with lamp, LCD, lens and ToF camera. b) Our prototype implementation uses a modified projector (instead of the* idealistic *LCD) and a depth camera.*

and amplitude at a modulation frequency of 20MHz. This allows for a depth range of 7.5m without ambiguities. Our ToF camera has a built-in illumination source, which operates in the infrared spectrum of 850 nm. The accuracy is highly dependent on the reflected amplitude as well as on the background illumination. According to [BOL*05] the best depth accuracy lies in the mm range. Our captured depth maps partially suffer from substantial noise, which has to be filtered as described in Sect. 4.1.

Digital SLR Camera

Our color camera is a Canon EOS 1D Mark III camera with a Canon EF 16-35mm f/2.8L USM lens. The camera is connected by USB to the PC. Using the Canon SDK we are able to control settings of the camera.

3.2. Calibration

The distortion, as well as the extrinsic and intrinsic matrices of the depth and the color camera, are computed from six images of the printed checkerboard pattern. This method can be used for both types of cameras, since the squares are also visible in the infrared spectrum of the depth camera. The projector is calibrated relative to the color camera based on the method presented in [SM00]. We use one printed colored checkerboard pattern for the projector calibration instead of black patterns. This allows us to capture only one instead of two images for the calibration. As we use a blue printed pattern and project a red pattern on the same plane, we can extract each of the two patterns independently without the need of touching the setup. We do this by simply considering one color channel at a time. Using a standard black pattern would force us to capture two images and cover the printed pattern manually when projecting onto the very same plane. Our approach therefore has the advantage, that the plane of reference will be exactly the same for both patterns, leading to a better accuracy, and more simplicity than the regular approach.

It is important to note that this calibration has to be done only once, assuming the SAFU is mounted rigidly to the camera.

4. Single-Shot Flash Adjustment

Our single-shot method permits compensation for the unnatural falloff of a conventional flash by virtually repositioning the light source farther back - behind the photographer. To simulate the falloff of such a light source, we require per-pixel depths. Our time-of-flight camera captures these depths at the same time the color camera is being aimed and focused. We call this method, where flash intensity is modulated based solely on depth, as "single-shot flash adjustment" (SSFA). SSFA has the advantage of minimizing motion artifacts relative to methods that require multiple color images. However, variations in object reflectance cannot be taken into account by this method.

Assuming an already calibrated setup, SSFA can be split into following steps:

1. Acquisition of depth and amplitude (infrared)
2. Depth filtering
3. Reprojection from the depth-camera to the flash unit
4. Per pixel, depth-dependent light adjustment
5. Color camera capture

The acquisition using the depth-camera is straight forward, and therefore, not elaborated any further in this paper.

4.1. Depth Filtering

The captured depth data has a direct influence on the projected intensity per pixel. Noise in the depth data is, therefore, also visible as noise in the evaluated illumination. Since

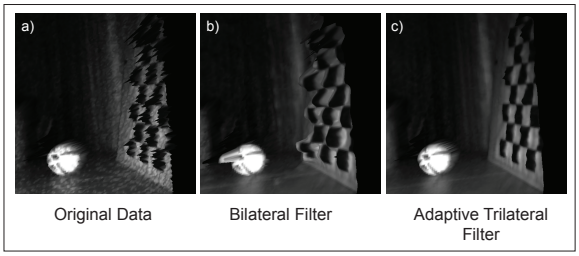


Figure 4: *a) Input depth data. b) Filtered data after applying* (5×5) *bilateral filter. c) Filtered data filtered with* (5×5) *adaptive trilateral filters.*

the human visual system is very sensitive to relative intensity variations, noise in the flash intensity becomes very easily visible. Although increasing the exposure time of the depth camera improves the SNR, it also increases motion blur artifacts. Therefore, we reduce the noise by applying different filters on the depth data allowing to keep a short exposure time.

Bilateral filtering [TM98] has shown to perform well for filtering noise while preserving discontinuities in the signal. However, in order to remove the noise, the discontinuity of the signal always has to be higher than the noise level, which is not always the case when dealing with depth data from ToF-cameras. Eisemann and Durand [ED04] and Petschnigg et al. [PSA*04] introduced the cross bilateral filter (a.k.a. joint bilateral filter), as a variant of the classical bilateral filter. They use a low noise image, meaning a high confidence in the high frequency data, as input to the range weight. This avoids smoothing across discontinuities of the signal. We extend this approach to a Joint Adaptive Trilateral Filter, where the noise model of the depth values of the ToF-camera is taken into account for the confidence of the range weights.

According to [BOL*05], the standard deviation σ_D of the depth data can be described as

$$\sigma_D = \frac{c}{4\pi \cdot f_{mod}\sqrt{2}} \cdot \frac{\sqrt{B}}{c_{demod} \cdot A_{sig}}, \text{ with} \quad (1)$$

$$c_{demod} = \frac{A}{B}, \text{ and} \quad (2)$$

$$B = A_{sig} + BG, \quad (3)$$

where c is the speed of light, f_{mod} is the modulation frequency, BG is the background illumination, A_{sig} is the mean number of electrons generated by the signal, and A is the measured amplitude in electrons. Additional noise sources, such as thermal noise or dark current electrons are minimized by the camera and are neglected by our filter. Since c and f_{mod} are constant and c_{demod} can be approximated by 0.5 according to [BOL*05], the standard deviation of the depth distribution is $\sigma_D \approx \frac{k}{\sqrt{A_{sig}}}$, with $k = \frac{c}{2\pi \cdot f_{mod}\sqrt{2}}$. This relationship shows that σ_D, and therefore, the noise in the depth image, mainly depends on the amplitude.

Joint Adaptive Trilateral Filter We introduce a novel joint adaptive trilateral filter (JATF), which takes the amplitude dependent standard deviation σ_D of the depth values into account. The JATF consists of a spatial proximity component $g(\cdot)$, a depth dependent component $f(\cdot)$ in the range domain and an amplitude dependent component $h(\cdot)$ in the range domain. The filtered depth value $d_{\mathbf{m}}$ at position $\mathbf{m}$ can be described as:

$$d_{\mathbf{m}} = \frac{1}{k_c} \sum_{\mathbf{i} \in \Omega} d_{\mathbf{i}} \cdot \underbrace{g(\|\mathbf{m} - \mathbf{i}\|)}_{spatial} \cdot \underbrace{f(d_{\mathbf{m}} - d_{\mathbf{i}}) \cdot h(a_{\mathbf{m}} - a_{\mathbf{i}})}_{range} \quad (4)$$

$$k_c = \sum_{\mathbf{i} \in \Omega} g(\|\mathbf{m} - \mathbf{i}\|) \cdot f(d_{\mathbf{m}} - d_{\mathbf{i}}) \cdot h(a_{\mathbf{m}} - a_{\mathbf{i}}) \quad (5)$$

where $a_{\mathbf{m}}$ and $a_{\mathbf{i}}$ are amplitude values at position $\mathbf{m}$ and $\mathbf{i}$ respectively. The three functions $g(\cdot)$,$f(\cdot)$ and $h(\cdot)$ are chosen to be Gaussian kernels, each with a kernel support of Ω, and the corresponding standard deviations σ_p, σ_d and σ_a respectively.

Intuitively, the filter is smoothing the depth values, while limiting the influence at depth or amplitude discontinuities. However, since the noise is not uniformly distributed, we adapt σ_d and σ_a, and therefore, the influence of $f(\cdot)$ and $h(\cdot)$. Having a good approximation of the uncertainty measure of the depth values Eq.(1), we can define two amplitude dependent functions $\sigma_d(\cdot)$ and $\sigma_a(\cdot)$ as follows:

$$\sigma_d(a_{\mathbf{m}}) = \sigma_{d_{init}} \cdot \left(\frac{\tau_a}{a_{\mathbf{m}}} + 1\right) \quad (6)$$

$$\sigma_a(a_{\mathbf{m}}) = \frac{\sigma_{a_{init}} - \sigma_{a_{min}}}{1 + e^{-a_{\mathbf{m}} + \tau_a}} + \sigma_{a_{min}}. \quad (7)$$

The initial standard deviation $\sigma_{d_{init}}$ is set to a low value, while the $\sigma_{a_{init}}$ is set to a high value, dropping to $\sigma_{a_{min}}$ at $a_{\mathbf{m}} = \tau_a$.

Median Filter For very low amplitudes, the provided depth is very often an outlier. Although the JATF would be able to remove such outliers, it would require several iterations to do so. Therefore, we apply a median filter, which removes the outliers in one step. To avoid unnecessarily blurring regions without outliers, we apply the median filter only to pixels with an amplitude below a threshold τ_m. A threshold of $\tau_m = 40$ has shown to be a good value.

4.2. Mesh-Based Reprojection

To attenuate the light according to the depth values, they have to be transformed from the depth camera view to the projector view. In order to solve this reprojection step efficiently, we form a trivial triangular mesh using all the back projected pixels from the depth camera and render them into the projector view. This provides a depth value per projector pixel and solves the visibility issue using the graphics card.

Artifacts can be introduced through resolution differences or remaining inaccuracies in the depth values. Since we do not have more information to improve the accuracy of the data

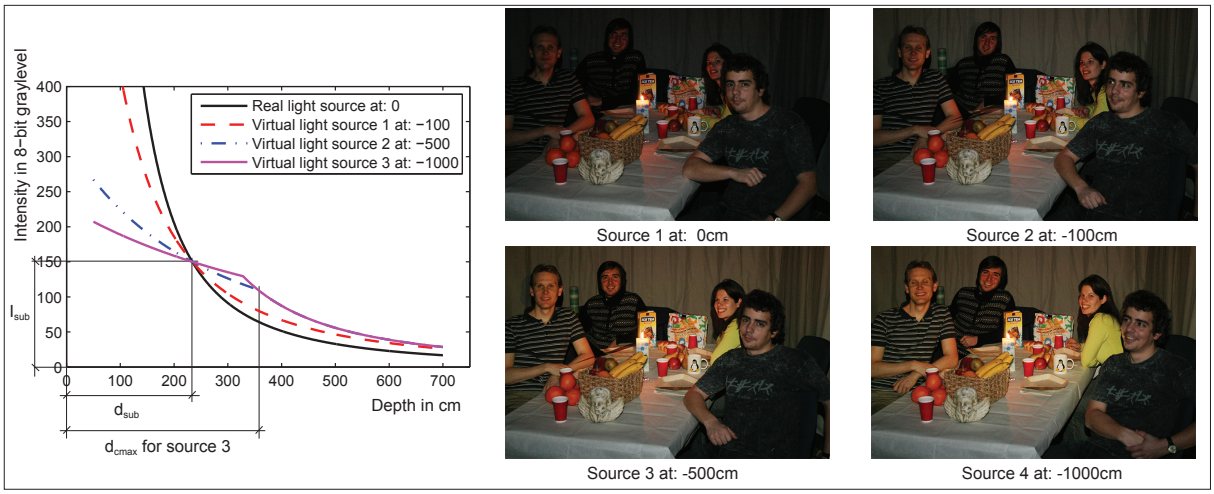


Figure 5: *In the four images we simulated virtual light sources positioned at different distances that are indicated in the captions and visualized in the plot. We observe that we cannot simulate source 3 beyond 380cm from the camera, because our flash unit offers insufficient power to stay on the blue curve. In practice, the transition point to the usual quadratic falloff (at d_{cmax}) is smooth, and we will not see artifacts.*

during the single-shot acquisition, we improve the quality of the result by moving the error to regions where it is least visible. Depth values that are too large produce visible artifacts, appearing as unexpectedly bright illumination. These artifacts are especially noticeable on objects close to the camera because of the steep falloff with increasing distance to the light source.

Similarly, depth values that are too small produce unexpectedly dim illumination. Most of these artifacts appear at depth discontinuities. This means, that dim illumination due to too small depth values, will be adjacent to shadows produced by the flash and thus be less noticeable than the artifacts caused by too large depth values. Therefore, we deliberately err on the side of computing depth values that are either correct or too shallow. We refer to this as the "conservative approach": a step we dub *shallow edge diffusion*.

Shallow Edge Diffusion For the shallow edge diffusion we start extracting all the large depth discontinuities using a Canny edge detector. Then, we simplify the obtained edges to a contour and move every vertex of it in the direction of its local gradient towards the region with the larger depth values. The amount of translation corresponds to the multiplication of a globally defined width and the local gradient of the vertex before translation. By connecting the first and last vertex of the original and the moved contour, we create a set $\mathbf{P}$ of points lying inside the polygon. Finally, we apply a modified bilateral filter (Eq.(8)) to all the depth values d_m laying inside of $\mathbf{P}$, such that the depth edge is filtered and being moved towards the region of bigger depth. The value k_c is normalizing the weights of the depth values d_i. $g(\cdot)$ and $h_p(\cdot)$ are two Gaussian kernels defined on the support $\Omega_\mathbf{P}$. $d_{\mathbf{i}_{min}}$ corresponds to the shallowest depth value of the kernel support.

$$d_\mathbf{m} = \frac{1}{k_c} \sum_{\mathbf{i} \in \Omega_\mathbf{P}} d_\mathbf{i} \cdot g(\|\mathbf{m} - \mathbf{i}\|) \cdot h_p(\|d_{\mathbf{i}_{min}} - d_\mathbf{i}\|) \qquad (8)$$

4.3. Depth Dependent Flash Adjustment

The intensity I_0 of an illumination source at distance 1 decreases with the distance d to an intensity of $I_d = \frac{I_0}{d^2}$. This intensity falloff becomes visible in flash-photography and leads to unnaturally lit scenes. A light source positioned farther behind the photographer, would result in a more homogeneously lit scene, which would appear more natural. Based on the depth values, we can evaluate a spatially dependent attenuation function, and simulate the more desirable light position as a virtual light source. The photographer can choose two interdependent parameters, namely the position of the virtual light source along the optical axis of the SAFU, and the minimal illumination for a specific subject of the scene. These preferences determine the depth compensation function leading to an attenuation per pixel.

The attenuation can in fact be interpreted as a filter by which the flash light is being modulated. In Sect. 4.3.3 we present two additional artistic filters, giving the photographer more creative freedom.

4.3.1. User Preferences

The parameters, which can be set by the user, are chosen based on the possible needs of the photographer. On the one hand, the photographer wants to pick the position of the light, and on the other hand, he wants to choose how much light is added to the subject of interest. The former can be set by providing the depth offset d_{off} relative to the

photographers position, with the positive axis lying in front of the photographer. The latter is defined by an intensity $I_{sub} \in \{0,1,2,..,255\}$ of the flash at the depth corresponding to the focal length d_{sub} of the camera. We think, that this is a natural way of parameterizing the virtual light source, since the photographer wants to primarily have control over how much light is added to the subject in focus. Especially for hand-held shots, the exposure time should not drop below a certain threshold to avoid motion blur due to camera shake. Of course, our method is not limited to this parametrization, but it is the one we thought to be most useful for candid shot settings.

4.3.2. Depth Compensation Function

Given the final depth values from Sect. 4.2 and the user parameters from Sect. 4.3.1, we can define the depth compensation function $f_c(\cdot)$ as

$$f_c(d_i) = \frac{I_{sub} \cdot (d_{sub} - d_{off})^2 \cdot d_{\mathbf{i}}^2}{(d_{\mathbf{i}} - d_{off})^2 \cdot d_{sub}^2} \tag{9}$$

$$f_{att}(d_i) = \begin{cases} I_{max} - f_c(d_i) & \text{if } f_c(d_i) < I_{max} \\ 0 & \text{if } f_c(d_i) \geq I_{max} \end{cases} \tag{10}$$

with $d_{\mathbf{i}}$ being the depth value at the pixel with index $\mathbf{i}$. $f_c(d_{\mathbf{i}})$ corresponds to the intensity that a pixel at index $\mathbf{i}$ requires to compensate for the position of the virtual light source. $f_{att}(\cdot)$ is the attenuation function dimming the maximal lamp power of the SAFU to result in a projection intensity of $f_c(\cdot)$. Since the compensation function is limited, we can define a maximal compensation depth d_{cmax} as

$$d_{cmax} = \frac{-I_{max} d_{sub} - \sqrt{I_{max} I_{sub} (d_{sub} - d_{off})^2} d_{off} d_{sub}}{I_{sub} d_{sub}^2 - 2 I_{sub} d_{sub} d_{off} + I_{sub} d_{off}^2 - I_{max} d_{sub}^2}, \tag{11}$$

where $I_{max} = f_c(d_{cmax})$. If d_{cmax} is smaller than any depth in the scene, the system could notify the photographer about a possible underexposure and either recommend a lower I_{sub} or a smaller offset d_{off}.

4.3.3. Artistic Filters

In addition to compensating for depth, one can also adapt the color balance to keep the lighting mood of the scene, such as in the candle lit scene of Fig. 6a. Furthermore, color images can be applied as artistic filters as well Fig. 6d.

5. Multi-Shot Flash Adjustment

The SSFA is limited by a low resolution depth image, as well as a low resolution infrared amplitude image. Therefore, scenes featuring very fine detail might show artifacts due to wrong depth values. Furthermore, the lack of color information does not allow to correct for varying reflectance properties and overexposed regions automatically. To improve on these two limitations, we implemented a "multi-shot flash adjustment" MSFA. Since a color image has to be captured before taking the final shot with the corrected flash light, we consider this as a multi-shot approach. Assuming an already calibrated setup as for the SSFA, the MSFA can be split into following steps:

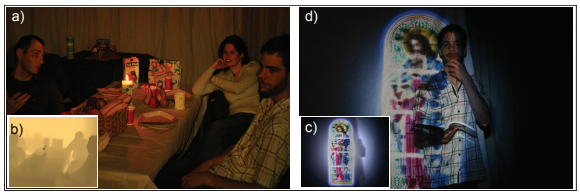


Figure 6: *a) A scene lit with a color filtered flash and the filter applied (inset). c) Artistic filter applied to scene d). d) A scene lit with a color image flash.*

1. Acquisition of depth and amplitude (infrared), and two color images
2. Depth filtering using high resolution color
3. Reprojection from the upsampled depth-data to the flash unit
4. Per pixel, depth-dependent light adjustment/reflectance correction
5. Color camera capture

Since most of the steps are very similar to the SSFA approach, we will focus on the enhanced filtering using color Sect. 5.1 and on the flash-tone-mapping Sect. 5.2.

5.1. Depth Filtering with Color

For reasons mentioned in Sect. 4.1, a sophisticated filtering is very important. Taking into account the high resolution color image, allows to filter and upsample the low resolution depth data, and therefore, get higher detail. Kopf et. al [KCLU07] presented the joint bilateral upsampling for images, where the mapping of corresponding areas between a low resolution image and a high resolution image is given. We map the depth data from the depth camera to the color camera by using the reprojection step described in Sect. 4.2. To improve the correspondences between the depth and the color values, we filter the depth data, according to Sect. 4.1, prior to the mapping. As a result we obtain the trivially upsampled depth data $\tilde{D}$ being of the same resolution as the color image $\tilde{I}$. To improve the performance of the filtering step, we apply a technique motivated by the joint bilateral upsampling [KCLU07].

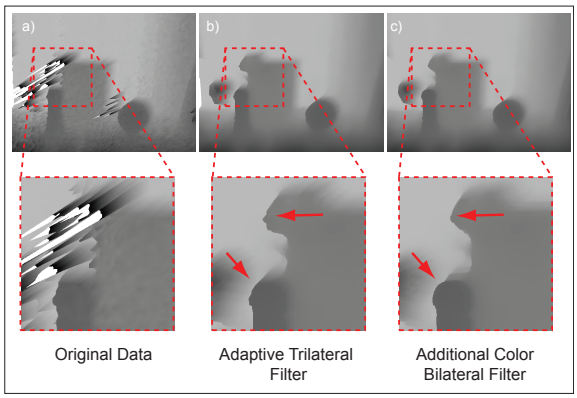


Figure 7: *a) Unfiltered depth contains noise and outliers. b) An adaptive trilateral filter reduces the noise, while preserving the depth discontinuities of the scene. Note the still present blocky edges due to the low resolution of the ToF camera. c) The modified joint bilateral upsampling based on the high resolution color camera refines the edges at depth discontinuities.*

$$\tilde{D}_{\mathbf{m}} = \frac{1}{k_c} \sum_{\mathbf{p}_{\downarrow\uparrow} \in \tilde{\Omega}_{\downarrow\uparrow}} \tilde{D}_{\mathbf{p}_{\downarrow\uparrow}} \cdot \tilde{g}(\|\mathbf{m} - \mathbf{p}_{\downarrow\uparrow}\|) \cdot \tilde{h_c}(\|\tilde{I}_{\mathbf{m}} - \tilde{I}_{\mathbf{p}_{\downarrow\uparrow}}\|) \tag{12}$$

$$k_c = \sum_{\mathbf{p}_{\downarrow\uparrow} \in \tilde{\Omega}_{\downarrow\uparrow}} \tilde{g}(\|\mathbf{m} - \mathbf{p}_{\downarrow\uparrow}\|) \cdot \tilde{h_c}(\|\tilde{I}_{\mathbf{m}} - \tilde{I}_{\mathbf{p}_{\downarrow\uparrow}}\|) \tag{13}$$

Instead of evaluating the full filter kernel, we evaluate every n^{th} pixel at position $\mathbf{p}_{\downarrow\uparrow}$ on the filter kernel $\Omega_{\downarrow\uparrow}$. n is the ratio between the width in pixels of the color camera, and the width in pixels of the depth camera, and for the height respectively. The closer the depth camera and the color camera are to a paraxial setting, the closer this approach is to the joint bilateral upsampling. The function $\tilde{g}(\cdot)$ and $\tilde{h_c}(\cdot)$ are Gaussian kernels with a kernel support of $\tilde{\Omega}$. The color at position m is referred to as $\tilde{I}_m$.

5.2. Flash-Tone-Mapping

In the previous sections we presented a method to adjust the amount of light per pixel depending on the per pixel depth value. However, we did not take the reflectance of the scene into account. Regions of the scene, which were not lit before adding the flash light, might saturate when doing so. Therefore, not taking into account the reflectance, would not allow to compensate for this. We propose to use a flash tone-mapping with a locally variable projection, which is motivated by the dodging-and-burning approach presented by Reinhard et. al in [RSSF02].

The response curve of the camera without flash $g(z_i)$ and the maximum extent of the response curve with flash $g_F(z_i)$, can

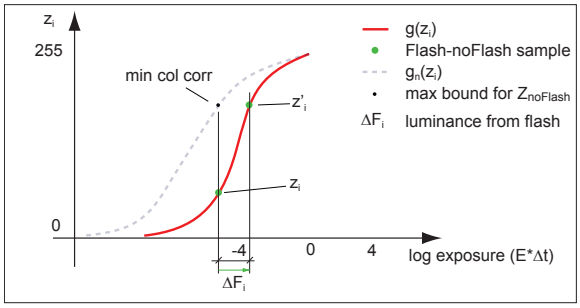


Figure 8: *The added luminance by the flash ΔF_i, can be evaluated from the difference of the log exposure $e^{g(z'_i)}$ of the flash image and the log exposure $e^{g(z_i)}$ of the no-flash image.*

be defined as:

$$\text{Camera:} \qquad g(z_i) = ln(\Delta t \cdot E_i) \tag{14}$$

$$\text{SAFU:} \qquad g_F(z_i) = ln(\Delta t \cdot E_i + \Delta F_i) \tag{15}$$

The image pixel values z_i are only dependent of the irradiance E_i and the chosen exposure time Δt for the camera response. $g(z_i)$ can be determined using a method presented by Debevec et. al [DM97]. Since the duration of the flash unit is usually much shorter than the exposure time, its response function $g_F(z_i)$ does not depend on the exposure time.

5.2.1. Depth Independent Range Extension

We want to find a new response curve $g_n(z_i)$ for the camera-flash system, which is similar to the response curve $g(z_i)$ of the camera, but extends its dynamic range independent of per pixel depth. Furthermore, it shall avoid overexposure in the flash lit image. $g_n(z_i)$ is bound by the lower bound $g_l(z_i)$ and the upper bound $g(z_i)$. The lower bound is determined as:

$$g_l(z_i) = ln(\Delta t \cdot E_i + \Delta F_{i_{max}}) \tag{16}$$

$$= ln\left(\Delta t \cdot E_i + \min\left(\Delta F_{xy}(z_i)\right)\right) \tag{17}$$

$$\Delta F_{xy}(z_i) = e^{g(I'_{xy})} - e^{g(I_{xy})}, \quad \forall xy \in I_{xy}. \tag{18}$$

The image pixels I_{xy} of the no-flash image at position (x,y), and the image pixels of the flash image I'_{xy} lie in the same range as z_i. The solution for $g_n(z_i)$ can now be approximated as a linear scaling of $g(z_i)$ by factor k (cf. Eq.(19)). $min(\frac{g_l(z_i)}{g(z_i)})$ complies with the lower bound $g_l(z_i)$. $g_n(z_i)$ could be smoothed slightly in order to avoid discontinuities in the fixed range (cf. Fig. 9).

Given $g_n(z_i)$ we can find the corresponding luminance contributed by the flash as

$$g_n(z_i) = k \cdot g(z_i) = min\left(\frac{g_l(z_i)}{g(z_i)}\right) \cdot g(z_i). \tag{19}$$

The scaling factor s_{xy} of the required light per pixel (x,y)

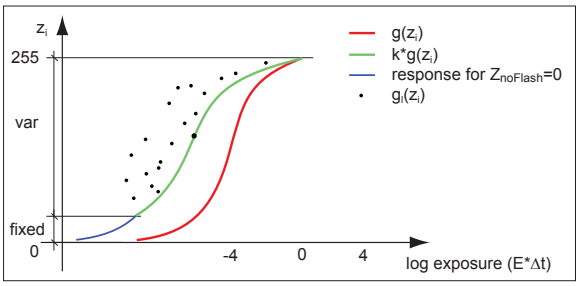


Figure 9: *The modified response function of the system has to lie in between the lower bound $g_l(z_i)$ and the upper bound $g(z_i)$. Due to the different depth in the scene, $g_l(z_i)$ is most often not a smooth curve. To illustrate one possible response function we choose to linearly stretch the camera response function respecting the lower range, while avoiding overexposure.*

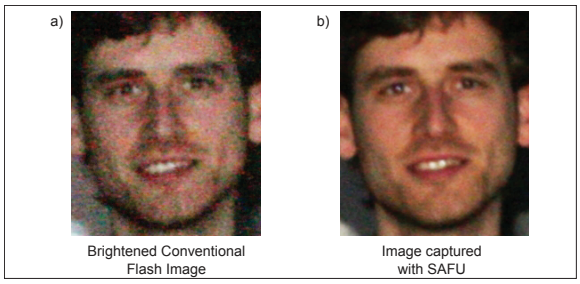


Figure 10: *a) Detail of a scene captured with a conventional flash and brightened up in a post-processing step. Noise as well as little detail in the color gradients become visible. b) Detail of a scene captured with the SAFU. The image shows very little noise.*

of the SAFU can now be evaluated as $s_{xy} = \frac{\Delta F_{i_{max}}}{\Delta F_{xy}}$ with $\Delta F_{i_{max}} = g_n(z_i) - g(z_i)$. Assuming only direct light contribution by the projector, lets us scale the maximum intensity of the corresponding projector pixel by the scaling factor s_{xy} as well.

6. Results

We built a prototype SAFU by replacing the light bulb of a SANYO PLC-XW50 projector with a Canon Speedlite 580EX II. The resolution of the projector is 1024×768 pixels. The depth is measured using a infrared based ToF-camera from Swissranger with a resolution of 176×144 pixels. The images were captured using a Canon 1D Mark III with a Canon EF 16-35mm f/2.8L USM lens. The resolution of the color camera is 3888×2592 pixels. We control the SAFU from our application, which is running under Windows XP on an Intel Core Duo with 2.6GHz and 3GB of RAM.

We captured different scenes using different settings in order to demonstrate the various capabilities of our method (Fig. 11, Fig. 12, and Fig. 13);

6.1. Discussion

We would like to compare our method to possible alternatives. They can be categorized into two different approaches. On the one hand, a captured image could be post-processed to achieve a seemingly even illumination. On the other hand, we could bounce the light off a ceiling to achieve an even illumination. We believe that the spatially adaptive photographic flash has advantages compared to either of them.

For instance, if an image is captured using a conventional flash and the distant parts are being amplified during post-processing, we also amplify the noise and lower the signal to noise ratio (SNR). The difference of the SNR of the most distant regions of the scene shown in Fig. 10a and the one in Fig. 10b is as high as 10dB, despite of the high quality camera we used in our setup. The standard deviation of the luminance is between 3 to 4 times higher for the conventional flash shown in Fig. 10a as compared to the SAFU in Fig. 10b (cf. Fig. 1(e) or Fig. 1(f)). For scenes with a bigger depth extent the difference of noise would increase as well. Furthermore, most consumer cameras have a smaller and lower quality CCD and show a much higher noise level than the camera applied in our setting, which strengthens the argument for the spatially adaptive illumination.

Instead of applying post-processing for an even illumination, one could also think of applying indirect lighting, by bouncing the flash light off a ceiling. Although, this method is often used in photography, it is strongly limited by the scene itself. In order to achieve an even but still bright illumination, the ceiling should be high enough but not too far away. Furthermore, it should be of a uniform and bright color, and should not contain any objects, which could cast an undesired shadow on the scene. These restrictions often make it rather difficult to apply bounced flash.

In summary, spatially adaptive photographic flash produces less noise than post-processing a uniformly illuminated scene, and it can be applied in situations where bounced flash is not practical.

7. Conclusion and Future Work

In this paper we have presented a spatially adaptive photographic flash, which improves on the limited illumination control of a common flash unit. By modulating flash intensity on a per-pixel basis, we can simulate a virtual light source further away from the subject than the physical light source. This reduces the harshness usually associated with flash illumination. We have built a prototype of our proposed flash unit, and we have shown the effect of moving the virtual light source to different positions. Furthermore, we have presented a flash tone-mapping technique that achieves an

(a) A scene with the virtual light source being placed 5m behind the photographer.

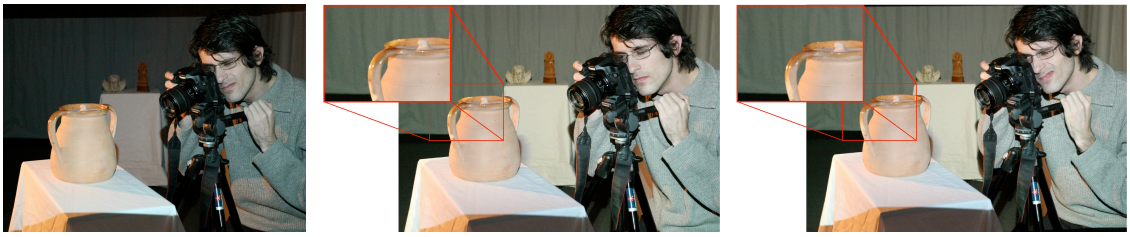
(b) Placing the light source 10m behind the photographer results in an almost homogeneous lighting of the scene.

(c) Applying a spatially adaptive flash without color correction to scene lit by a candle removes the nice warm tones from the candle light.

(d) Adding a color filter keeps the mood of the scene, while still brightening it up for a shorter exposure time.

(e) More artistic effects can be created with an image filter, such as a *virtual* window.

Figure 11: *Mood preservation using colored masks (11(a)–11(c)) & Artristic filters (11(e)).*



(a) Scenes with a big depth extent result in a harsh lighting when photographed with a regular flash.

(b) The single-shot allows to compensate for the intensity falloff, but suffers from the low resolution of the ToF-camera, which can result in misalignments.

(c) We take advantage of the high resolution of the color camera, to improve the accuracy of our spatial compensation.

Figure 12: *SAFU multi-shot corrects artifacts.*

(a) This scene is only lit by a desklamp in a dark room resulting in a wide dynamic range.

(b) shows the same scene as in Fig. 13(a), but captured with a full intensity flash burst.

(c) The same scene as in Fig. 13(a), but with added flash light in the dark regions, without over-exposing the bright ones.

Figure 13: *SAFU tone mapping.*

effect similar to "dodging and burning". Our spatially adaptive photographic flash relies on depth data to compute its spatial modulation. To reduce noise in this data, we have presented a novel joint adaptive trilateral filter, which adapts to the noise model of the depth data. While we believe our method has potential, it also suffers from several limitations.

7.1. General Limitations

One of the most crucial limitations of our prototype setup is the efficiency of the maximal light output. Using an LCD-projector for the spatial light modulation, drops the light by at least 50%, even if the attenuation of the SAFU is set to zero. This is due to the limited possibility to align the flash unit to the light path inside the projector. It is currently limiting the application of the SAFU to low light settings. Unfortunately it seems that a substitution by DMD technology, as has successfully been realized for light modulation during acquisition [NBB06], is not an option for a flash unit. The non-controllable duration of the flash burst is in the order of 1.2ms, and therefore, much shorter than the required integration time of the DMD. The limitation of the maximal intensity as well as the limited contrast are setting constraints on the maximal depth compensation the SAFU can deal with. However, dedicated hardware like the one presented in Fig. 3a) would remove the loss of light caused by the sub-optimal light path through the projector.

Our system contains only one SAFU, such that we cannot compensate for shadows. By adding a second unit to the system, these shadows could be minimized.

7.2. Single-Shot

Depth Camera The single-shot method is strongly based on the data provided by the ToF-camera. Non-IR reflecting objects or very specular objects, such as mirror-like surfaces, cannot reliably be reconstructed. Furthermore, the limited resolution can become visible as aliased edges at sharp depth-discontinuities.
Processing Time The current processing time lies at 0.064s for a single-shot acquisition. This processing time limits the amount of motion which can be present in the scene. The motion does not appear as motion blur in the images, but as a misalignment of the geometry with the flash compensation function. We believe, however, that our algorithm could be implemented directly on a processing unit of modern digital cameras. By doing so the processing time would drop substantially.
Spatial Limitation The spatial limitation is twofold. On the one hand, the ToF-camera features a range of 7.5m without any depth ambiguities, and on the other hand, the limited light efficiency of our *prototype* allows for a compensation of a couple of meters depending on the settings.

7.3. Multi-Shot

Color Camera Acquisition The time to acquire a color image over the USB connection using the Canon SDK lies in the range of one second. This is a severe limitation and makes it impossible to capture spontaneous shots. This is, however, a limitation of our specific setup and could be removed if the algorithm was implemented directly on the camera or in the flash unit.
Limited Dynamic Range The current SAFU implementation does not show a very high light efficiency. This has a direct impact on the amount of correction we can apply to a scene.

7.4. Future Work

In our work we use a spatially modulated flash mainly to compensate for deficiencies in a scene's natural illumination. However, structured illumination can accomplish many other things, such as separating direct and indirect illumination [NKGR06] or changing the apparent color, shape, or surface properties of objects [RWLB01]. By introducing additional projectors or optics, one can create a 4D illumination light field, which can be used to estimate the surface properties of objects [DHT*00], digitize their shape [RHHL02, ZCHS03], eliminate foreground or background objects [LCV*04], or even compute an image as seen from the flash's point of view [SCG*05].

References

[ARNL05] AGRAWAL A., RASKAR R., NAYAR S. K., LI Y.: Removing photography artifacts using gradient projection and flash-exposure sampling. In *SIGGRAPH '05: ACM SIGGRAPH 2005 Papers* (New York, NY, USA, 2005), ACM, pp. 828–835.

[BOL*05] BÜTTGEN B., OGGIER T., LEHMANN M., KAUFMANN R., LUSTENBERGER F.: CCD/CMOS lock-in pixel for range imaging: Challenges, limitations and state-of-the-art. In *First Range Imaging Research Day* (2005), pp. 21–23.

[DHT*00] DEBEVEC P., HAWKINS T., TCHOU C., DUIKER H.-P., SAROKIN W., SAGAR M.: Acquiring the reflectance field of a human face. In *SIGGRAPH '00: Proceedings of the 27th annual conference on Computer graphics and interactive techniques* (New York, NY, USA, 2000), ACM Press/Addison-Wesley Publishing Co., pp. 145–156.

[DM97] DEBEVEC P. E., MALIK J.: Recovering high dynamic range radiance maps from photographs. In *SIGGRAPH '97: Proceedings of the 24th annual conference on Computer graphics and interactive techniques* (New York, NY, USA, 1997), ACM Press/Addison-Wesley Publishing Co., pp. 369–378.

[ED04] EISEMANN E., DURAND F.: Flash photography enhancement via intrinsic relighting. In *ACM Transactions on Graphics* (Aug 2004), vol. 23, pp. 673–678.

[KCLU07] KOPF J., COHEN M. F., LISCHINSKI D., UYTTENDAELE M.: Joint bilateral upsampling. *ACM Transactions on Graphics 26*, 3 (2007), 96.

[LCV*04] LEVOY M., CHEN B., VAISH V., HOROWITZ M., MCDOWALL I., BOLAS M.: Synthetic aperture confocal imaging. *ACM Transactions on Graphics 23*, 3 (2004), 825–834.

[MBW*07] MOHAN A., BAILEY R. J., WAITE J., TUMBLIN J., GRIMM C., BODENHEIMER B.: Tabletop computed lighting for practical digital photography. *EEE Transactions on Visualization and Computer Graphics 13*, 4 (2007), 652–662.

[NB03] NAYAR S. K., BRANZOI V.: Adaptive dynamic range imaging: Optical control of pixel exposures over space and time. *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on 2* (Oct. 2003), 1168–1175.

[NBB06] NAYAR S. K., BRANZOI V., BOULT T. E.: Programmable Imaging: Towards a Flexible Camera. *International Journal on Computer Vision 70*, 1 (Oct 2006), 7–22.

[NKGR06] NAYAR S. K., KRISHNAN G., GROSSBERG M. D., RASKAR R.: Fast separation of direct and global components of a scene using high frequency illumination. *ACM Transactions on Graphics 25*, 3 (2006), 935–944.

[PSA*04] PETSCHNIGG G., SZELISKI R., AGRAWALA M., COHEN M., HOPPE H., TOYAMA K.: Digital photography with flash and no-flash image pairs. *ACM Transactions on Graphics 23*, 3 (2004), 664–672.

[RHHL02] RUSINKIEWICZ S., HALL-HOLT O., LEVOY M.: Real-time 3d model acquisition. In *SIGGRAPH '02: Proceedings of the 29th annual conference on Computer graphics and interactive techniques* (New York, NY, USA, 2002), ACM, pp. 438–446.

[RSSF02] REINHARD E., STARK M., SHIRLEY P., FERWERDA J.: Photographic tone reproduction for digital images. *ACM Transactions on Graphics 21*, 3 (July 2002), 267–276.

[RWLB01] RASKAR R., WELCH G., LOW K.-L., BANDYOPADHYAY D.: Shader lamps: Animating real objects with image-based illumination. In *Proceedings of the 12th Eurographics Workshop on Rendering Techniques* (London, UK, 2001), Springer-Verlag, pp. 89–102.

[SCG*05] SEN P., CHEN B., GARG G., MARSCHNER S. R., HOROWITZ M., LEVOY M., LENSCH H. P. A.: Dual photography. *ACM Transactions on Graphics 24*, 3 (2005), 745–755.

[SM00] SHEN T.-S., MENQ C.-H.: Digital projector calibration for 3d active vision systems. *Transactions of the ASME 124 124* (2000), 126–134.

[SNB07] SCHECHNER Y. Y., NAYAR S. K., BELHUMEUR P. N.: Multiplexing for optimal lighting. *IEEE Transactions on Pattern Analysis and Machine Intelligence 29*, 8 (2007), 1339–1354.

[TM98] TOMASI C., MANDUCHI R.: Bilateral filtering for gray and color images. In *ICCV '98: Proceedings of the Sixth International Conference on Computer Vision* (Washington, DC, USA, 1998), IEEE Computer Society, p. 839.

[ZCHS03] ZHANG L., CURLESS B., HERTZMANN A., SEITZ S. M.: Shape and motion under varying illumination: Unifying structure from motion, photometric stereo, and multi-view stereo. In *ICCV '03: Proceedings of the Ninth IEEE International Conference on Computer Vision* (Washington, DC, USA, 2003), IEEE Computer Society, p. 618.