Flash Photography Enhancement via Intrinsic Relighting
Elmar Eisemann, Frédo Durand

To cite this version:

HAL Id: inria-00510164
https://inria.hal.science/inria-00510164
Submitted on 13 Oct 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Flash Photography Enhancement via Intrinsic Relighting

Elmar Eisemann∗
MIT / ARTIS-GRAVIR/IMAG-INRIA

Frédéric Durand
MIT

Abstract

Our technique enhances photographs shot in dark environments by combining a picture taken with the available light and one taken with the flash. We preserve the ambiance of the original lighting and insert the sharpness. We use the bilateral filter to decompose the images into detail and large scale. We reconstruct the image using the large scale of the available lighting and the detail of the flash. We detect and correct flash shadow. Our output combines the advantages of available illumination and flash photography.

Keywords: Computational photography, flash photography, re-lighting, tone mapping, bilateral filtering, image fusion

1 Introduction

Under dark illumination, a photographer is usually faced with a frustrating dilemma: to use the flash or not. A picture relying on the available light usually has a warm atmosphere, but suffers from noise and blur (Fig. 1(a) top and (b)). On the other hand, flash photography causes three unacceptable artifacts: red eyes, flat and harsh lighting, and distracting sharp shadows at silhouettes (Fig. 1(a) bottom). While much work has addressed red-eye removal [Zhang and Lenders 2000; Gaubatz and Ulichney 2002], the harsh lighting and shadows remain a major impediment.

We propose to combine the best of the two lightings by taking two successive photographs: one with the available lighting only, and one with the flash. We then recombine the two pictures and take advantage of the main qualities of each one (Fig. 1(c)). Our central tool is a decomposition of an image into a large-scale layer that is assumed to contain the variation due to illumination, and a small-scale layer containing albedo variations.

Related work

Most work on flash photography has focused on red-eye removal [Zhang and Lenders 2000; Gaubatz and Ulichney 2002]. Many cameras use a pre-flash to prevent red eyes. Professional photographers rely on off-centered flash and indirect lighting to prevent harsh lighting and silhouette shadows.

Our work is related to the continuous flash by Hoppe and Toyama [2003]. They use a flash and a no-flash picture and combine them linearly. The image-stack interface by Cohen et al. [2003] provides additional control and the user can spatially vary the blending. Raskar et al. [2004] and Akers et al. [2] fuse images taken with different illumination to enhance context and legibility. DiCarlo et al. [2001] use a flash and a no-Flash photograph for white balance.

Multiple-exposure photography allows for high-dynamic-range images [Mann and Picard 1995;Debevec and Malik 1997]. New techniques also compensate for motion between frames [Ward 2004; Kang et al. 2003]. Note that multiple-exposure techniques are different from our flash-photography approach. They operate on the same lighting in all pictures and invert a non-linear and clamped.

Figure 1: (a) Top: Photograph taken in a dark environment, the image is noisy and/or blurry. Bottom: Flash photography provides a sharp but flat image with distracting shadows at the silhouette of objects. (b) Inset showing the noise of the available-light image. (c) Our technique merges the two images to transfer the ambiance of the available lighting. Note the shadow of the candle on the table.
response. In contrast, we have a quite-different lighting in the two images and try to extract the lighting ambiance from the no-flash picture and combine it with the fine detail of the flash picture.

We build on local tone-mapping techniques that decompose an image into two or more layers that correspond to small- and large-scale variations, e.g. [Chiu et al. 1993; Jobson et al. 1997; Tumblin and Turk 1999; DiCarlo and Wandell 2000; Durand and Dorsey 2002; Reinhard et al. 2002; Ashikhmin 2002; ?]. Only the contrast of the large scales is reduced, thereby preserving detail.

These methods can be interpreted in terms of intrinsic images [Tumblin et al. 1999; Barrow and Tenenbaum 1978]. The large scale can be seen as an estimate of illumination, while the detail corresponds to albedo [Oh et al. 2001]. Although this type of decoupling is hard [Barrow and Tenenbaum 1978; Weiss 2001; Tappen et al. 2003], tone mapping can get away with a coarse approximation because the layers are eventually recomposed. We exploit the same approach to decompose our flash and no-flash images.

A wealth of efforts has been dedicated to relighting, e.g. [Marschner and Greenberg 1997; Sato et al. 1999; Yu et al. 1999]. Most methods use acquired geometry or a large set of input images. In contrast, we perform lighting transfer from only two images.

In this volume, Petschnigg et al. [?] present a set of techniques based on flash/no-flash image pairs. Their decoupling approach shares many similarities with our work, in particular the use of the bilateral filter. The main difference between the two approaches lies in the treatment of flash shadows.

### 2 Image decoupling for flash relighting

Our approach is summarized in Fig. 2. We take two photos, with and without the flash. We align the two images to compensate for camera motion between the snapshots. We detect the shadows cast by the flash and correct the color using local white balance. We finally perform a non-linear decomposition of the two images into large-scale and detail layers and we recombine them appropriately.

We first present our basic technique before discussing shadow correction in Section 3. We then introduce more advanced reconstruction options in Section 4 and present our results in Section 5.

**Taking the photographs** The two photographs with and without the flash should be taken as rapidly as possible to avoid motion of either the photographer or subject. The response curve between the two exposures should ideally be known for better relative radiometric calibration, but this is not a strict requirement. Similarly, we obtain better results when the white balance can be set to manual. In the future, we foresee that taking the two images in a row will be implemented in the firmware of the camera. To perform our experiments, we have used a tripod and a remote control (Fig. 1 and 8) and hand-held shots (Fig. 2, 5, 7). The latter in particular requires good image alignment. In the rest of this paper, we assume that the images are normalized so that the flash image is in $[0, 1]$.

The registration of the two images is not trivial because the lighting conditions are dramatically different. Following Kang et al. [2003], we compare the image gradients rather than the pixel values. We use a low-pass filter with a small variance (2 pixels) to smooth-out the noise. We keep only the 5% highest gradients and we reject gradients in regions that are too dark and where information is not reliable. We use a pyramidal refinement strategy similar to Ward [2004] to find the transformation that minimizes the gradients that were kept. More advanced approaches could be used to compensate for subject motion, e.g. [Kang et al. 2003].

**Bilateral decoupling** We first decompose each image into layers corresponding to the illumination and the sharp detail respectively. We use the bilateral filter [Tomasi and Manduchi 1998; Smith and Brady 1997] that smoothes an image but respects sharp features, thereby avoiding halos around strong edges [Durand and Dorsey 2002].

The bilateral filter is defined as a weighted average where the weights depend on a Gaussian $f$ on the spatial location, but also on a weight $g$ on the pixel difference. Given an input image $I$, the output of the bilateral filter for a pixel $s$ is:

$$J_s = \frac{1}{k(s)} \sum_{p \in \Omega} f(|p - s|) g(I_p - I_s) I_p,$$  \hspace{1cm} (1)

where $k(s)$ is a normalization: $k(s) = \sum_{p \in \Omega} f(|p - s|) g(I_p - I_s)$. In practice, $g$ is a Gaussian that penalizes pixels across edges that have large intensity differences. This filter was used by Oh et al. [2001] for image editing and by Durand et al. for tone mapping [2002].

We use the fast bilateral filter where the non-linear filter is approximated by a set of convolutions [Durand and Dorsey 2002]. We perform computation in the $\log_{10}$ domain to respect intensity ratios. The output of the filter provides the log of the large-scale layer. The detail layer is deduced by a division of the intensity by the large-scale layer (subtraction in the log domain). We use a spatial variance $\sigma_s$ of 1.5% of the images diagonal. For the intensity influence $g$, we use $\sigma_g = 0.4$, following Durand and Dorsey [2002].
Reconstruction

Ignoring the issue of shadows for now, we can recombine the image (Fig. 2). We use the detail and color layer of the flash image because it is sharper and because white balance is more reliable. We use the large-scale layer of the no-flash picture in order to preserve the mood and tonal modeling of the original lighting situation. The layers are simply added in the log domain. Fig. 3 illustrates the results from our basic approach. The output combines the sharpness of the flash image with the tonal modeling and texture contrast of the no-flash image.

For dark scenes, the contrast of the large scale needs to be enhanced. This is the opposite of contrast reduction [Durand and Dorsey 2002]. We set a target contrast for the large-scale layer and scale the range of log values accordingly. The low quantization from the original image does not create artifacts because the bilateral filter results in a piecewise-smooth large-scale layer.

In addition, we compute the white balance between the two images by computing the weighted average of the three channels with stronger weights for bright pixels with a white color in the flash image. We then take the ratios $w_f$, $w_p$, $w_b$ as white-balance coefficients. This white balance can be used to preserve the warm tones of the available light. In practice, the color cast of the no-flash image is usually too strong and we only apply it partially using $w^i$ where $i$ is usually 0.2.

We must still improve the output in the flash shadow. While their intensity is increased to match the large scale of the no-flash image, there is a distinct color cast and noise. This is because, by definition, these areas did not receive light from the flash and inherit from the artifacts of the no-flash image. A ring flash might reduce these artifacts, but for most cameras, we must perform additional processing to alleviate them.

3 Shadow treatment

In order to correct the aforementioned artifacts, we must detect the pixels that lie in shadow. Pixels in the umbra and penumbra have different characteristics and require different treatments. After detection, we correct color and noise in the shadows. The correction applied in shadow is robust to false positives; Potential detection errors at shadow boundaries do not create visible artifacts.

Umbra detection

We expect the difference image $\Delta I$ between flash and no-flash to tell how much additional light was received from the flash. When the images are radiometrically calibrated, $\Delta I$ is exactly the light received from the flash. However, shadows do not always correspond to $\Delta I = 0$ because of indirect lighting. While shadow pixels always correspond to the lowest values of $\Delta I$, the exact cutoff is scene-dependent.

We use histogram analysis to compute a threshold $t_{\Delta I}$ that determines umbra pixels. Shadows correspond to a well-marked mode in the histogram of $\Delta I$. While the additional light received by parts of the scene lit by the flash varies with albedo, distance and normal, the parts in shadow are only indirectly illuminated and receive a more uniform and very low amount of light.

We compute the histogram of pixels $\Delta I$. We use 128 bins and smooth it with a Gaussian blur of variance two bins. We start with a coarse threshold of 0.2 and discard all pixels where $\Delta I$ is above this value. We then use the first local minimum of the histogram before 0.2 as our threshold for shadows detection (Fig. 5). This successfully detects pixels in the umbra. However, pixels in the penumbra correspond to a smoother gradation and cannot be detected with our histogram technique. This is why we use a complementary detection based on the gradient at shadow boundaries.

Penumbra detection

Shadow boundaries create strong gradients in the flash image that do not correspond to gradients in the no-flash image. We detect these pixels using two criteria: the gradients difference and connectedness to umbra pixels.

We compute the magnitude of the gradient $V\Delta I$ and $V\Delta f$ and smooth it with a Gaussian of variance 2 pixels to remove noise. We identify candidate penumbra pixels as pixels where the gradient is stronger in the flash image. We then keep only pixels that are “close” to umbra pixels, that is, such that at least one of their neighbors is in umbra. In practice, we use a square neighborhood of size 1% of the photo’s diagonal. This computation can be performed efficiently by convolving the binary umbra map with a box filter.

We also must account for shadows cast by tiny objects such as...
The wealth of information provided by the pair of images can be exploited to enhance results for very dark situations and more advanced lighting transfer.

4 Advanced decoupling

The wealth of information provided by the pair of images can be further exploited to enhance results for very dark situations and more advanced lighting transfer.
Figure 7: The flash lighting results in a flat image. In our result, light seems to be coming from the window to the right.

received conflicting feedback on Fig. 7, which shows that image quality is a subjective question. In this image, the light is coming from the 3/4 back, which is an unusual lighting for a photograph. Some viewers appreciate the strong sense of light it provides, while others object to the lack of tonal modeling.

Another cause of failure is overexposure of the flash, leading to a flat detail layer. In this situation, the detail information is neither in the no-flash (due to noise) nor in the flash photo (due to saturation). Shadow detection works best when the depth range is limited. Distant objects do not receive light from the flash and are detected in shadow. While this is technically correct, this kind of shadow due to falloff does not necessitate the same treatment as cast shadow. Fortunately, our color correction is robust to false positives and degrades to identity in these cases (although transition areas could potentially create problems). Similarly, black objects can be detected as shadows, but this does not affect quality since they are black in the two images and remain black in the output. Light flares can cause artifacts by brightening shadow pixels. The method by Ward [2004] could alleviate this problem.

We have used our algorithms with images from a variety of cameras including a Sony Mavica MVC-CD400 (Fig. 1), a Nikon Coolpix 4500 (all other images), a Nikon D1 and a Kodak DC4800 (not shown in the paper). The choice of the camera was usually dictated by availability at the time of the shot. The specifications that affected our approach are the noise level, the flexibility of control, the accuracy of flash white balance, and compression quality. For example, the Kodak DC4800 exhibited strong JPEG artifacts for dark images, which required the use of the cross-bilateral filter.

The need for the cross-bilateral filter was primarily driven by the SNR in the no-flash picture. The Kodak DC4800 has higher noise levels because it is old. Despite its age, the size of its photosites allows the Nikon D1 to take images in dark conditions. In addition, the use of the RAW format with 12 bits/channel allows for higher precision in the flash image (the lower bits of the no-flash image are dominated by noise). However, with the sensitivity at 1600 equivalent ISO, structured noise makes cross-bilateral filtering necessary.

6 Conclusions and future work

We have presented a method that improves the lighting and ambiance of flash photography by combining a picture taken with the flash and one using the available lighting. Using a feature-preserving filter, we estimate what can be seen as intrinsic layers of the image and use them to transfer the available illumination to the flash picture. We detect shadows cast by the flash and correct their color balance and noise level. Even when the no-flash picture is extremely noisy, our method successfully transfers lighting due to the use of the flash image to perform edge-preserving filtering.

The method could be tailored to particular cameras by fine-tuning parameters such as $\sigma_g$ based on a sensor-noise model. Additional red-eye removal could benefit from the additional information provided by the pair of images. Texture synthesis and inpainting could be used to further improve shadow correction. Ideally, we want to alleviate the disturbance of the flash and we are considering the use of infrared illumination. This is however challenging because it requires different sensors and these wavelengths provide limited resolution and color information.

The difference of the flash and no-flash images contains much information about the 3D scene. Although a fundamental ambiguity remains between albedo, distance and normal direction, this additional information could greatly expand the range and power of picture enhancement such as tone mapping, super-resolution, photo editing, and image based-modeling.

Acknowledgments We acknowledge support from an NSF CISE Research Infrastructure Award (EIA-9802220) and a Desh-
Appendix: Intensity-Color decoupling. Traditional approaches rely on linear weighted combinations of $R$, $G$, and $B$ for intensity estimation. While these formulae are valid from a color-theory point of view, they can be improved for illumination-albedo decoupling. Under the same illumination, a linear intensity computation results in lower values for primary-color albedo (in particular blue) than for white objects. As a result, the intensity transfer might overcompensate as shown in Fig. 9(left) where the red fur becomes too bright. To alleviate this, we use the channels themselves as weights in the linear combination:

$$ I = \frac{R}{R+G+B} + \frac{G}{R+G+B} + \frac{B}{R+G+B} $$

In practice, we use the channels of the flash image as weight for both pictures to ensure consistency between the two decoupling operations. The formula can also be used with tone mapping operators for higher color fidelity.

Figure 9: The computation of intensity from RGB can greatly affect the final image. Left: with linear weights, the red pixels of the fur become too bright. Right: using our non-linear formula.

©ACM, 2004. This is the author’s version of the work. It is posted here by permission of ACM for your personal use. Not for redistribution. The definitive version was published in Proceedings of SIGGRAPH 2004 (ACM Transactions on Graphics (TOG)), archive Volume 23, Issue 3 (August 2004)

References


