REMOVING SHADOWS USING FLASH/NOFLASH IMAGE EDGES

Mark S. Drew[1]*, Cheng Lu*[1]*, and Graham D. Finlayson*[2]

[1]School of Computing Science, Simon Fraser University, Vancouver, Canada V5A 1S6
[2]School of Computing Sciences, University of East Anglia, Norwich, U.K. NR4 7TJ
{mark,clu}@cs.sfu.ca, graham@cmp.uea.ac.uk

ABSTRACT

Flash/noflash pairs have been used for noise-reduction in ambient-light images. But not explicitly studied is the problem of shadows in the ambient images. While shadows are lessened in a flash image, other problems arise, and other shadows are produced. It is known that we can in fact produce a flash-only (no ambient) image by subtracting the two images, but the result is not as pleasant as the ambient image, because of several artifacts due to the flash. Here, we use the pure-flash image to detect the ambient shadows. We argue that first going to a "spectrally sharpened" color space, and then focusing on the difference in a log domain of the flash image minus the ambient image, gives a very simple feature space consisting of two components — one in an illuminant-change 3-vector direction, and one along the gray axis. This space provides excellent separation of the shadow and nonshadow areas. Inserting edges from the flash image within the ambient-shadow region into the ambient image edge map and inverting Poisson's equation fills in the shadow. In this way, we arrive at an image with the advantages of the ambient-only image — warmth, no flash effects such as disturbing illumination dropoff with distance, pixel saturation etc. — but no shadows.

1. INTRODUCTION

The combination of both ambient lighting plus a flash of course produces quite different illumination at a surface point than in an image taken under ambient lighting. For clarity, let us refer to the first image as "Ambient" and the second as "Both" (A and B). If we control the camera settings, or at least know them, and assuming there are no saturated pixels etc. [1], then $(B-A)$ should yield an image as if it were taken under the flash only (assuming one adjusts overall pixel magnitudes to compensate for camera settings, as in [2]). This is due to the fact that the B image consists of reflected light from the ambient sources plus from the flash. Lighting is from a different direction from the camera flash than the effective direction for ambient lighting (at each pixel). That is, there is a different visibility function for the flash image, and this produces "flash-shadows". But since the pure-flash image, denoted "Flash" (F), sees reflected light from only the flash illumination, image F has no ambient shadows — no shadows that derive from the ambient, usually quite visible in both B as well as in A.

So we should be able to combine F with A, the ambient image, to be able to detect and hence eliminate the ambient-shadows. But we found that a simple differencing scheme does not work well: $F - A$ is large within the ambient-shadow region, but this is confounded by the change in color of the two illuminants: ambient and flash illuminations. Here, we re-cast the difference image $F - A$ to make ambient-shadows easily detected. We go over to a color space in which a simple, diagonal, model of illumination change applies; then the difference of log images forms a feature space with two very simple components: a constant component along a 3-vector lighting-change direction, plus a difference of log intensity/shading/visibility terms along the gray $(1, 1, 1)$ axis. A projection into this plane leaves a large intensity difference between in-shadow pixels from A and from F, compared to out-of-shadow pixels. Results are seen to be very effective: ambient-shadows are simply identified.

Work on flash/noflash pairs started with [3] and then [2, 4, 5]. These works do not explicitly use the flash information for finding shadows (finding a "shadow matte", in graphics terms–cf. [6]). Instead, these efforts are aimed at using the high-frequency information and detail from the lower-noise flash+ambient image to reduce noise in the ambient-only image, e.g. using a joint bilateral filter. But the flash image by itself suffers from several drawbacks that make using the ambient-only image appealing: illumination dropoff with distance (producing a "tunnel effect" [7]), pixel saturation, and strong interreflections. The ambient image is known to be warmer and more pleasing. Related work has also been carried out on filling in information in nighttime imagery using much brighter, daytime images [8, 9], or using multiple flashes [10, 11], but again not for eliminating ambient shadows.

Here we show that a simplified image model greatly aids in finding the shadows in the ambient image: naively using the intensity cannot succeed, since dark pixels can arise from many sources. But if we consider image formation, for the flash and ambient images, we can find a color space representation that separates the ambient-shadow feature vectors from the out-of-shadow locations. Our approach to removing ambient-shadows uses the high-quality gradient information from the flash image for the shadow regions and retains the nonshadow regions in the ambient image. To seamlessly combine these two parts, the image is reconstructed by solving Poisson equations.

In §2, a simple image-formation model is described that greatly simplifies the problem. Section 3 sets out how to promote the simple model via a color-space transform, and then examines the log-difference image for flash-noflash pairs, showing that an illumination difference image can be generated with no surface-color term confounding lighting change. Section 4 recovers the shadow mask for the ambient image, via projecting log-difference values onto the plane they mostly occupy and then detecting large differences by

a robust procedure. Then §5 shows how to copy flash information over to the ambient image, inside the shadow mask, and then re-integrate by combining gradients to guarantee integrability. The result is an ambient image with much less ambient shadowing.

2. IMAGE FORMATION

Let us restrict attention to Lambertian surfaces. Then at a surface point, under orthography, lighting is added up into a single effective light [12], taking into account visibility factors for each source. In general, the direction of the effective light is different in each color band [13].

Consider the RGB $\boldsymbol{R}$ formed at a pixel $\boldsymbol{x}$, for illumination $E(\lambda)$ lighting a surface with reflectance function $S(\lambda)$. With 3 camera sensor sensitivity functions $\boldsymbol{Q}\ (\lambda)$, we have

$$R_k \ = \ \sigma \int E(\lambda)S(\lambda)Q_k(\lambda)d\lambda \ , \ k = R, G, B \ , \quad (1)$$

where σ is Lambertian shading — surface normal dotted into illumination direction — along with visibility.

We wish to go to a model that explains the difference in images formed under different lights by a simple diagonal 3×3 matrix. It has been found that this illuminant-change model holds, and in fact the image-formation description is greatly simplified, if we make assumptions of Planckian lighting, Lambertian surfaces, and a narrowband camera [14]. In this case, we find that a log-difference image for flash F and ambient A images has a very simple form.

With a Dirac delta camera sensor $Q_k(\lambda) = q_k\delta(\lambda-\lambda_k)$, eq. (1) becomes simply

$$R_k \ = \ \sigma \ E(\lambda_k)S(\lambda_k)q_k \ . \quad (2)$$

Approximating lighting by Planck's law, in Wien's approximation:

$$E(\lambda, T) \quad \simeq \ I\, k_1\lambda^{-5}e^{-\frac{k_2}{T\lambda}} \ , \quad (3)$$

with constants k_1 and k_2, temperature T characterizes the lighting color and I gives the overall light intensity. In this approximation, from (2) the RGB color R_k, $k = 1 \ldots 3$, is simply given by

$$R_k \ = \ \sigma \ I \ k_1\lambda_k^{-5}e^{-\frac{k_2}{T\lambda_k}}S(\lambda_k)q_k \ . \quad (4)$$

Define the following short-hand notations:

$$\begin{array}{lll} K \ = \ \log(Ik_1\sigma); & s_k \ = \ \log(S(\lambda)); \\ w_k \ = \ \log(k_1\lambda_k^{-5}q_k); & e_k \ = \ -k_2/\lambda_k \end{array} \quad (5)$$

Taking logarithms, eq. (4) becomes

$$\log R_k(\boldsymbol{x}\ ) = w_k \ + \ K(\boldsymbol{x}\ ) \ + \ s_k(\boldsymbol{x}\ ) \ + \ (1/T(\boldsymbol{x}\ ))e_k \quad (6)$$

Here, the 3-vector w_k depends on the camera, as does e_k, independent of image location. The intensity and shading, encapsulated in K, do depend on location, as does the surface term s_k. Lighting temperature T depends on what lighting the surface point sees and adds up. *In shadows, both K and T are different* — the shadow region sees a different color and intensity of light than does the nonshadowed area. In this paper, we assume that lighting is effectively independent of location $\boldsymbol{x}$ in each separate region: shadow and nonshadow.

Using these simplifying assumptions, we will use the fact that the ambient shadow is removed from the pure-flash image F to find the ambient-shadows in A.

3. NONNEGATIVE SPECTRAL SHARPENING AND IMAGE DIFFERENCE MODEL

3.1. Spectral Sharpening

The simplified model (4) is more closely followed if $\boldsymbol{Q}\ (\lambda)$ approximates a Dirac delta. We form an intermediate color space in which the sensors are optimally combined so as to form new colors that better approximate color change induced by illuminant change via a diagonal model using Spectral Sharpening [15]. This applies a 3×3 transformation matrix $\boldsymbol{M}$ to the sensors, or directly to colors, so as to better enforce a diagonal model.

Since we mean to take logs, we need nonnegative colors from the camera data (with zero values treated specially). To do so, we carry out a "spectral sharpening with positivity" transform (cf. [16, 17]). Using calibration targets under two different lights, We find $\boldsymbol{M}$ via a novel optimization, new to this paper, consisting of a constrained form of "database sharpening" [15], with hard constraints:

$$\begin{array}{l} \underset{\boldsymbol{M}\ ,\boldsymbol{D}}{min} \ \ \sum[\boldsymbol{M}\ \boldsymbol{D}\ \boldsymbol{M}^{\ -1}A \ - \ B]^2 \\ \text{with constraints} \begin{cases} -\boldsymbol{M}\ A \ < \ 0 \\ \text{non-negative sensor result,} \\ \\ \boldsymbol{D} \ = \ \text{diagonal matrix} \end{cases} \end{array} \quad (7)$$

3.2. Log Image Difference

From (6), we notice that a difference image in log space space can remove both a camera term w_k as well as the surface term $s_k(\boldsymbol{x}\ )$. Let us form a *ratio image* by a difference image in log space: subtracting eq. (6) for two images, A and F, we have

$$\begin{array}{l} \log R_k^A(\boldsymbol{x}\ ) - \log R_k^F(\boldsymbol{x}\ ) \ = \ [K^A(\boldsymbol{x}\ ) - K^F(\boldsymbol{x}\ )] \\ + \ [1/T^A(\boldsymbol{x}\ ) - 1/T^F(\boldsymbol{x}\ )]e_k \end{array} \quad (8)$$

for the difference between log pixel values under light A and light F, at pixel indexed by $\boldsymbol{x}$. Notice that the surface term is entirely removed, leaving a type of *intrinsic illumination* difference image. In this image, shadowing will stand out.

The ratio image arises from: (i) the intensity difference (with shading/visibility), proportional to the basic direction $\boldsymbol{u} \ = 1/\sqrt{3}(1, 1, 1)^T$; (ii) a term proportional to the camera-dependent lighting-change 3-vector e_k. Notice that eq. (8) is pixel-wise: the two parts would be different inside a shadow, if one is present, since the lighting color temperature would be different there than in the rest of the image, as would the intensities. For details, including plots showing that eq. (8) actually obtains in real imagery, please see [18]. We shall use image Fig. 1 to illustrate the method in the following.

4. SHADOW-MATTE ALGORITHM

To find ambient-shadows, we need to find a plane in which most (non-shadow) pixels live, with basis vectors given by the $(1, 1, 1)$ direction $\boldsymbol{u}$ and the lighting-change direction $\boldsymbol{e}$, as in eq. (8).

It is easy to find this plane, using SVD. Then we can best model the log-difference by projecting onto this plane:

$$\boldsymbol{d} \ = \ \boldsymbol{V}\ (\log \boldsymbol{R}^{\ A} - \log \boldsymbol{R}^{\ F}) \quad (9)$$

where $\boldsymbol{V}$ is the 2×3 pair of principal component vectors in the plane: $\boldsymbol{d}$ is 2-dimensional. Projecting discards some information, but mostly discards noise.

We found that most of the variation in log-difference magnitude is along the $\boldsymbol{u}$ direction (in the sharpened color

space). Therefore we apprehend shadow formation as a process that generates large values in the log difference between Flash and Ambient (i.e., the $\boldsymbol{u}$ direction), as well as differences in lighting color (the $\boldsymbol{e}$ direction). To take both into account we simply use the Euclidean magnitude in the plane. Since the logarithm is monotonic, and we are seeking large differences, this will find outlier values from both causes.

We use a robust method [19] (Least Median of Squares–LMS) to find flash-shadows and specularities from $\|F\|$, using a 1D location mode-finder. The LMS gives outliers automatically, and these are shown in Fig. 2(a) and In general, outliers will be either in flash-shadows, or in specularities, with outliers on the dark side of the mode and specularities on the bright side. We include both kinds of outliers in a "flash-shadow/specularity" mask since we wish to exclude them both.

Then we derive a magnitude $\mathcal{I}$ in the plane as $\mathcal{I} = \|\boldsymbol{d}\|$. Again we use the LMS, and the resulting ambient-shadow mask (excluding the flash-shadows) is shown in Fig. 2(b). As an auxiliary step, we additionally produce a cleaner shadow mask by applying a Mean-Shift segmentation [20] to the $\mathcal{I}$ image, before finding outliers.

5. SHADOW-FREE AMBIENT IMAGE RECOVERY

Finding the ambient-shadow mask takes us most of the way toward recovery of a *shadowless* ambient image. To proceed, we first map flash-image pixels to pixels in the ambient image, inside the shadow mask. To do so, we first regress on pixel values not in either the ambient shadow or flash shadow/specularities, from F to A. Then we use the resulting 3×3 matrix $\boldsymbol{T}$ to take flash pixels over to ambient ones. Then a straightforward approach would join the regressed F(in ambient shadow) and A(out of ambient shadow).

However, this does not entirely eliminate the ambient shadows. This is because (i) the regression does not give perfect color mapping from F to A; and (ii) A has illumination discontinuities that do not usually exactly coincide with the mask boundary. But we may note that *edges* are not discontinuous in each image F and A inside the ambient-shadow mask. Therefore, to seamlessly join the two parts, we can blend the image edges from F and A by combining their gradient fields: if S is the shadow mask, then we define a new field $\log \boldsymbol{C}$ via

$$\begin{array}{rcl} \nabla \log \boldsymbol{C}[S] & = & \nabla(\log \boldsymbol{F}[S]\boldsymbol{T}) \\ \nabla \log \boldsymbol{C}[\neg S] & = & \nabla(\log \boldsymbol{A}[\neg S]) \end{array} \qquad (10)$$

Taking another gradient and inverting Poisson's equation to recover $\log \boldsymbol{C}$ will blend the two gradient fields. We can simply use a Fourier transform on the entire image to re-integrate, a very fast method, using homogeneous Neumann boundary conditions. As well, in order to combine the Ambient and Flash gradients such that integrability is preserved, we project the combined set of edges onto an integrable convex set in the Fourier domain, before re-integrating via inverse Fourier transform [21, 22]. The re-integrated ambient image for our example image is shown in Fig. 2(c); the 3D plot of pixels in feature space shwon in Fig. 2(d). Additional results are shown in [18].

Altogether, the algorithm for ambient-shadow removal is as follows:

Algorithm:

Find sharpening matrix $\boldsymbol{M}$ from camera calibration.
Transform to sharpened color space:
$A^\sharp = \boldsymbol{M}\ A$; $B^\sharp = \boldsymbol{M}\ B$, $F^\sharp = B^\sharp\ -\ A^\sharp$.
Form logs, and log-difference $\log F^\sharp - \log A^\sharp$.
Find best 2D plane for log-difference feature: basis $\boldsymbol{V}$.
Project log-difference into 2D $\boldsymbol{d}$: eq. (9).
Find flash-shadow/specularity mask S_F=
outliers LMS($\|F\|$).
Apply mean-shift filtering to $\|\boldsymbol{d}\|$.
Find ambient-shadow mask S_A= outliers LMS($\|\boldsymbol{d}\|$).
Combine masks: $S = S_A \cap (\neg S_F)$.
Copy color from A into F, to set scale for flash image edge map:
$\boldsymbol{P}$ =LS($F[\neg(S_A \cup S_F)], A[\neg(S_A \cup S_F)]$),
where LS is least-squares regression; then $F' = \boldsymbol{P}\ F$.
Dilate S.
Form combined log-image edge map $\log \boldsymbol{C}$ via eq. (10).
In each color channel, we effectively take another derivative in the Fourier domain via a phase shift, and then project onto an integrable edge map whilst undoing the Laplacian.
The inverted Poisson equation is unique only up to an additive constant, which we fix by regressing back to the ambient image $\log A$, in $(\neg S)$.
Exponentiate to recover non-log image.

---

6. CONCLUSION

We present a method of removing shadows from ambient images for ambient/ambient+flash pairs. The method relies on determining an illumination-field image from the log-difference of the pure flash and ambient images. This image removes the surface-reflectance, yielding intrinsic illumination and shading information. In an ambient shadow region, the difference between the two is large, from an intensity difference or a lighting-color difference or both, so finding the shadow area is straightforward. Copying flash-image information over to the ambient image, flash and ambient information is blended such that the resulting integrable edge field generates a shadowless, ambient-lighting, result.

References

[1] R. Ramanath, W.E. Snyder, Y.F. Yoo, and Mark S. Drew, "Color image processing pipeline in digital still cameras," *IEEE Signal Processing*, vol. 22, no. 1, pp. 34–43, 2005.

[2] G. Petschnigg, R. Szeliski, M. Agrawala, M. Cohen, H. Hoppe, and K. Toyama, "Digital photography with flash and no-flash image pairs," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 664–672, 2004.

[3] J. M. DiCarlo, F. Xiao, and B. A. Wandell, "Illuminating illumination," in *Color Imaging Conf.*, 2001, pp. 27–34.

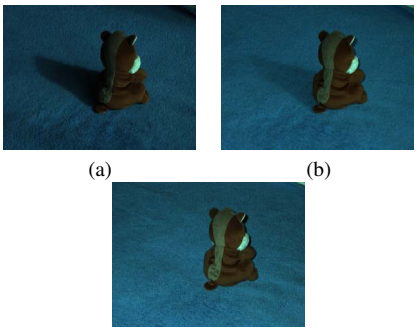
[4] E. Eisemann and F. Durand, "Flash photography enhancement via intrinsic relighting," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 673–678, 2004.

[5] A. Agrawal, R. Raskar, S. K. Nayar, and Y. Li, "Removing photography artifacts using gradient projection and flash-exposure sampling," *ACM Trans. Graph.*, vol. 24, no. 3, pp. 828–835, 2005.

[6] Y.-Y. Chuang, D. B Goldman, B. Curless, D. H. Salesin, and R. Szeliski, "Shadow matting and compositing," *ACM Trans. Graph.*, vol. 22, pp. 494–500, 2003.

[7] A. Agrawal, R. Raskar, S. K. Nayar, , and Y. Li, "Removing photography artifacts using gradient projection and flash-exposure sampling," *ACM Trans. Graph.*, vol. 24, pp. 828–835, 2005.

[8] R. Raskar, A. Ilie, and J. Yu, "Image fusion for context enhancement and video surrealism," in *NPAR '04: 3rd*



(a) (b) (c)

Fig. 1. Sample images: (a): Ambient, (b): Both, and (c): Flash. These images are 12-bit per channel linear consumer digital color camera images, in raw format with no gamma or other processing applied. However, images are shown in the sRGB color space, for display [23].

Int. Symp. on Non-photorealistic Animation and Rendering, 2004, pp. 85–152.

[9] R. Raskar, A. Ilie, and J. Yu, "Gradient domain context enhancement for fixed cameras," in *ACCV 2004*, 2004, pp. 27–30.

[10] J. J. Yoon, C.Koch, and T. J. Ellis, "ShadowFlash: an approach for shadow removal in an active illumination environment," in *BMVC*, 2002, pp. 636–645.

[11] R. Raskar, K. T., R. Feris, J. Yu, and M. Turk, "Non-photorealistic camera: Depth edge detection and stylized rendering using multi-Flash imaging," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 679–688, 2004.

[12] W. Silver, *Determining Shape and Reflectance Using Multiple Images*, Ph.D. thesis, MIT, 1980.

[13] M.S. Drew, "Robust specularity detection from a single multi–illuminant color image," *CVGIP: Image Understanding*, vol. 59, pp. 320–327, 1994.

[14] G.D. Finlayson and S.D. Hordley, "Colour constancy at a pixel," *J. Opt. Soc. Am. A*, vol. 18, no. 2, pp. 253–264, Feb. 2001.

[15] G.D. Finlayson, M.S. Drew, and B.V. Funt, "Spectral sharpening: sensor transformations for improved color constancy," *J. Opt. Soc. Am. A*, vol. 11, no. 5, pp. 1553–1563, May 1994.

[16] M.S. Drew and G.D. Finlayson, "Spectral sharpening with positivity," *J. Opt. Soc. Am. A*, vol. 17, pp. 1361–1370, 2000.

[17] K. Barnard, F. Ciurea, , and B. Funt, "Sensor sharpening for computational color constancy," *J. Opt. Soc. Am. A*, vol. 18, pp. 2728–2743, 2001.

[18] M.S. Drew, C. Lu, and G.D. Finlayson, "Ambient shadow detection and removal via flash/noflash pairs," Tech. Rep., Simon Fraser University, 2006, http://fas.sfu.ca/pub/cs/TR/2006/CMPT2006-05.pdf.

[19] P. J. Rousseeuw and A. M. Leroy, *Robust Regression and Outlier Detection*, Wiley, 1987.

[20] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *PAMI*, vol. 24, pp. 603–619, 2002.

[21] T. Simchony, R. Chellappa, and M. Shao, "Direct analytical methods for solving Poisson equations in computer vision problems," *IEEE Trans. Patt. Anal. and Mach. Intell.*, vol. 12, pp. 435–445, 1990.

[22] G.D. Finlayson, S.D. Hordley, C. Lu, and M.S. Drew, "On the removal of shadows from images," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 28, pp. 59–68, 2006.

[23] HP/Microsoft, ," http://www.srgb.com.

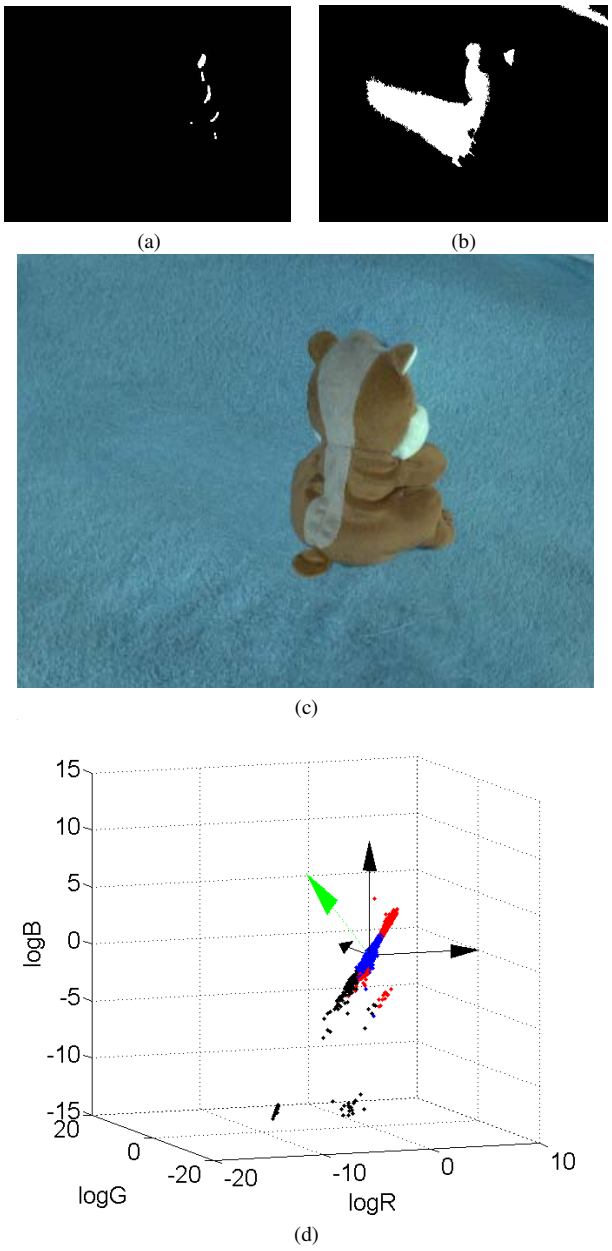


Fig. 2. (a): Flash-shadow mask. (b): Shadow mask for ambient shadows. (c): Re-integrated image without ambient or flash shadows. (d): Pixels identified as ambient-shadow (red) and flash-shadow (black); the coordinate axes are also shown, along with the perpendicular to the plane which the data inhabit (green vector).