Dark Flash Photography

Dilip Krishnan* Rob Fergus

Dept. of Computer Science, Courant Institute, New York University

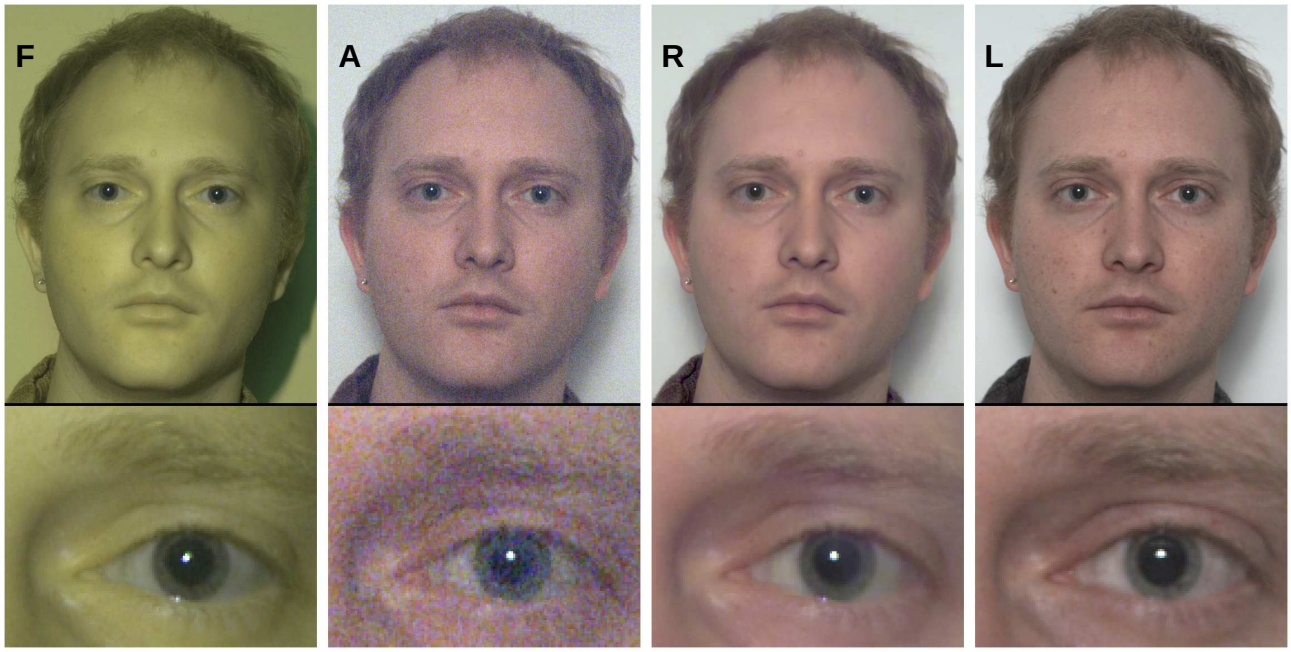


Figure 1: *Our camera and flash system offers dazzle-free photography by hiding the flash in the non-visible spectrum. A pair of images are captured at a blur-free shutter speed, one using a multi-spectral flash (F), the other using ambient illumination (A) which in this case is 1/100th of that required for a correct exposure. The pair are combined to give an output image (R) which is of comparable quality to a reference long exposure shot (L). The figures in this paper are best viewed on screen, rather than in print.*

Abstract

Camera flashes produce intrusive bursts of light that disturb or dazzle. We present a prototype camera and flash that uses infra-red and ultra-violet light mostly outside the visible range to capture pictures in low-light conditions. This "dark" flash is at least two orders of magnitude dimmer than conventional flashes for a comparable exposure. Building on ideas from flash/no-flash photography, we capture a pair of images, one using the dark flash, other using the dim ambient illumination alone. We then exploit the correlations between images recorded at different wavelengths to denoise the ambient image and restore fine details to give a high quality result, even in very weak illumination. The processing techniques can also be used to denoise images captured with conventional cameras.

Keywords: Computational Photography, Dark Flash, Multi-Spectral Imaging, Spectral Image Correlations

*Email: dilip@cs.nyu.edu

1 Introduction

The introduction of digital camera sensors has transformed photography, permitting new levels of control and flexibility over the imaging process. Coupled with cheap computation, this has precipitated a wide range of novel photographic techniques, collectively known as Computational Photography. Modern camera sensors, be they in a cellphone or a high-end DSLR, use either a CCD or CMOS sensor based on silicon. The raw sensor material responds to light over a wide range of wavelengths, typically 350–1200nm. Colored dyes are deposited onto the sensor pixels in a Bayer pattern, resulting in 3 groups of pixels (red, green and blue). Each responds to a limited range of wavelengths, approximating the sensitivities of the three types of cone cell in our retina. However, silicon is highly sensitive to infra-red (IR) wavelengths and it is difficult to manufacture dyes that have sufficient attenuation in this region, thus an extra filter is placed on top of most sensors to block IR light. This gives a sensor that records only over the range 400-700nm, matching our own color perception, but a considerable restriction of the intrinsic range of the device.

One solution to capturing photographs in low light conditions is to use a flash unit to add light to the scene. Although it provides the light to capture otherwise unrecordable scenes, the flash makes the photographic process intrusive. The sudden burst of light not only alters the illumination but disturbs any people present, making them aware that a photo has just been taken and possibly dazzling them if they happen to be looking toward the camera. For example, a group photo in a dark restaurant or bar using a bright camera flash leaves the subjects unable to see clearly for some moments afterward.

In this paper we introduce a camera/flash system that is based around off-the-shelf consumer equipment, with a number of minor modifications. First, the camera is a standard DSLR with the IR-block filter removed, thus restoring much of the original spectral range of the sensor. Second, we use a modified flash that emits light over a wider spectral range than normal, which we filter to remove visible wavelengths. This *dark flash* allows us to add light to the scene in such a way that it can be recorded by the camera, but not by our own visual system. Using the dark flash we can illuminate a dimly lit scene without dazzling people present, or significantly disturbing those around. Furthermore, it allows a fast shutter speed to be used, thus avoiding camera shake. However, the difficulty is that people want images with colors that match their visual experience and this will not be the case for images captured using the dark flash.

To overcome this, we acquire a pair of images in the manner of flash/no-flash photography [Eisemann and Durand 2004; Petschnigg et al. 2004], one using the dark flash and the second using ambient illumination alone. For the latter to be blur-free a fast shutter speed must be used, resulting in high noise levels in dim light. A key observation is that if the non-visible and visible channels are close in wavelength, strong correlations will exist between them. We introduce a novel type of constraint that exploits the correlations between spectral bands. Using this constraint, the edge structure of the dark flash image can be used to remove the noise from the ambient image, yielding a high quality result that lacks the shadow and specularity artifacts present in the flash image.

We also show how our camera/flash hardware and spectral constraints can be used in a range of additional applications, including: inferring spectral reflectance functions of materials in the scene and denoising individual color channels of images captured with standard cameras.

1.1 Related work

Our approach can be regarded as a multi-spectral version of the flash/no-flash technique introduced by [Agrawal et al. 2005], [Petschnigg et al. 2004] and [Eisemann and Durand 2004]. [Agrawal et al. 2005] focused on the removal of flash artifacts but did not apply their method to ambient images containing significant noise, unlike [Petschnigg et al. 2004] and [Eisemann and Durand 2004]. The two latter approaches are similar in that they use a cross-bilateral (also known as joint-bilateral) filter and detail transfer. However, [Petschnigg et al. 2004] attempt to denoise the ambient, adding detail from the flash, while [Eisemann and Durand 2004] alter the flash image using ambient tones.

The closest work to ours is that of [Bennett et al. 2007], who show how video captured in low-light conditions can be denoised using continuous IR illumination. However, they make use of temporal smoothing to achieve high quality results, something not possible in our photography setting. [Wang et al. 2008a] show how IR illumination can be used to relight faces in well-lit scenes. Both these works differ from ours in a number of ways: (i) they use complex optical bench based setups with twin cameras and beam-splitters – we use a single portable DSLR camera and temporally multiplex instead; (ii) both use IR alone rather than the near-UV and IR that we use (both being necessary for high quality reconstructions); (iii) both rely on cross-bilateral filtering to combine the IR and visible signals, an approach which we demonstrate to have serious shortcomings. In contrast, we propose a principled mechanism for propagating information between spectral bands. We integrate this into a unified cost function that combines the denoising and detail transfer mechanisms, treated separately in cross-bilateral filtering and related methods, such as [Farbman et al. 2008].

Infra-red imaging has a long history in areas such as astronomy and night-vision. In consumer photography the most prominent use has been the Sony Nightshot where the IR-block filter can be switched out to use the near-IR part of the spectrum. The images are monochrome (with a greenish tint) and no attempt is made to restore natural colors to them. Other imaging approaches use Far-IR wavelengths to record the thermal signature of people or vehicles. However, this requires specialized optics and sensors and thus has limited relevance to consumer photography. Ultra-violet (UV) photography has received little attention, other than from flower photography enthusiasts [Rorslett 2008]. Many flowers that look plain to humans have vibrant patterns under UV light to attract insects sensitive to these wavelengths.

Multi-spectral recording using visible wavelengths has been explored by several authors. [Park et al. 2007] used multiplexed illumination via arrays of colored LEDs to recover spectral reflectance functions of the scene at video frame rates. Our system can be used in a similar manner for still scenes, being able to estimate the reflectance functions beyond the visible range. [Mohan et al. 2008] use a diffraction grating in conjunction with an LCD mask to give control over the color spectrum for applications including metamer detection and adaptive color primaries.

Our processing of the flash/no-flash pair exploits the correlations *between* nearby spectral bands. Most work on image priors has focused on capturing spatial correlations *within* a band. For example, priors based on the heavy tailed distributions of image gradients have proven highly effective in a wide range of problems such as denoising [Portilla et al. 2003], deblurring [Fergus et al. 2006], separating reflections [Levin and Weiss 2007]. However, models that exploit dependencies between color channels are less common. The K-SVD denoising approach of [Aharon et al. 2006] does so implicitly by vector quantizing color patches. The fields-of-experts approach of [Roth and Black 2005] has also been extended to model color images [McAuley et al. 2006] and uses color marginal filters. However, neither of these approaches explicitly model the inter-channel correlations, unlike our method. Explicit spectral models are used in color constancy problems and joint spatial-spectral models have been proposed [Singh et al. 2003; Chakrabarti et al. 2008] for this task, but these assume a noise-free image. [Morris et al. 2007] measured the spatial gradients of far IR images gathered with a specialized camera, demonstrating their similarity to those of visible light images.

Flash-based methods are not the only solution to taking pictures in low-light levels. Wide aperture lenses gather more light but are heavy and expensive, making them impractical for most photographers. Anti-shake hardware can be used to capture blur-free images at slow shutter speeds. These techniques can be combined with our approach to extend performance to even lower light levels. Software-based deblurring techniques [Fergus et al. 2006; Jiaya 2007] can only cope with modest levels of blur and typically have artifacts in their output. Denoising techniques [Tomasi and Manduchi 1998; Portilla et al. 2003] have similar performance issues, and cannot cope with the noise levels we address in this paper. Joint denoising/deblurring techniques, such as that of Yuan [Yuan et al. 2007], provide better performance but still require a problematic deconvolution operation, which can introduce artifacts. Methods that register and combine a stack of noisy images, such as [Telleen et al. 2007], have the inconvenience of needing to capture far more than two images. Finally, a visible flash can be made non-dazzling by using a diffuser and aiming at the ceiling. This methods works well but is limited to indoors settings with a fairly low ceiling of neutral color.

2 Dark flash hardware

In our approach we capture a pair of images, one with the dark flash (F) and another using ambient lighting alone (A). The pixel

value p in channel j of image F depends on three terms: the spectral response of each camera channel $C_j(\lambda)$ at wavelength λ; the illumination spectrum of the dark flash $I^f(\lambda)$; and the surface reflectance function $S(p,\lambda)$ at the point in the scene. These combine in a linear fashion:

$$F_j(p) = \int C_j(\lambda) I^f(\lambda) S(p,\lambda)\ d\lambda \qquad (1)$$

with $j = \{1,2,3\}$ being the index of the camera channel. Note we assume even illumination (i.e. $I^f(\lambda)$ does not depend on p). The ambient image A is formed in a similar fashion, using illumination $I^a(\lambda)$ which scales with the exposure interval. A_1, A_2 and A_3 record red, green and blue wavelengths respectively under typical illumination. Through the choice of flash and camera, we can control $I^f(\lambda)$ and the channel sensitivities $C_j(\lambda)$.

A primary design constraint is that off-the-shelf consumer hardware should be used where possible, making the system cheap and easily reproducible. Our camera is a Fuji IS Pro, which is marketed for applications involving UV and IR work since it lacks an IR sensor filter. The flash is a Nikon SB-14UV. Both the camera and the flash are equipped with carefully chosen filters, detailed in Appendix A, that shape both $I^f(\lambda)$ and $C_j(\lambda)$ for our application. These filters remain in place for both shots, thus the pair of images can be taken in quick succession, limited only by the 3 frames/sec rate of the camera. The flash is used at full power for all shots, the cycle time being sufficiently long that it does not fire for the second shot, giving an image with ambient illumination alone. The system is no more complex to operate than a standard DSLR (see Fig. 3(top left) for a picture of the system).

We now describe the form of $I^f(\lambda)$ and how it can be recorded by the camera while remaining largely invisible to humans. The spectral response of each camera channel $C_j(\lambda)$ is shown in Fig. 2(a). Note that with no IR sensor filter, the responses extend considerably beyond the visible range (400-700nm). The spectrum of the dark flash $I^f(\lambda)$ is shown in Fig. 2(b). It has two distinct emission lobes, both just outside the visible range. The first, consisting of UV light, couples with the small part of channel $j = 3$'s response extending below 400nm. The second lobe in the IR region between 700 and 800nm is picked up by channel $j = 1$ which responds strongly. Thus, the dark flash allows the recording of two independent measurements at each location in a scene within a single image: one in UV recorded in F_3, the other in IR recorded in F_1.

The flash/no-flash image pair captures the scene at 5 different spectral bands, assuming the ambient illumination is dim compared to the output of the flash: 1. UV (370–400nm) in F_3; 2. Blue ($\sim$ 400–500nm) in A_3; 3. Green ($\sim$ 500–600nm) in A_2; 4. Red ($\sim$ 600-700nm) in A_1 and 5. IR (700nm–800nm), recorded in F_1. In Fig. 3, we show a Macbeth color chart in each of these five bands.

For comparison purposes, we also use a standard visible flash whose power is adjusted to give comparable camera exposure to the dark flash. In Fig. 3(top) we attempt to show the relative perceived brightness of the dark and visible flashes by capturing them using a standard DSLR whose spectral response is close to that of our eyes (thus the brightness in the image should correspond to our perception). See Section 4.3 for a quantitative analysis of their relative brightness.

Safety issues. As shown in Fig. 2(b), our dark flash emits energy just outside visible wavelengths, centered around 380nm with negligible energy below 360nm or above 400nm (until the IR lobe at 700nm). The health hazard posed by UV light depends strongly on the wavelength, those close to visible (400nm) being orders of magnitude safer than the shorter wavelength components of sunlight. Our flash is very close to visible, even closer than blacklights found in bars and nightclubs, which have a broader spectral width centered at 360nm. In the USA, the acknowledged regulations regarding the safe daily exposure to UV light are given in the Threshold Limit Values (TLV) booklet, published by the government body ACGIH [TLVs 2001]. We carefully measured the absolute spectral irradiance of our flash using a spectrometer. Using the TLV tables, the maximum safe number of flashes per day can be computed, which is 130,000 at 1m from the flash. Put another way, if we assume that 30 minutes outside in the sun results in the maximum permissible UV dose on a bright summer day, then each flash is equivalent to being outside for 1/100th second. Hence our dark flash poses no significant safety hazard. Details of these calculations can be found in Appendix B.

3 Dark flash processing

The pair of images, F and A are captured using a shutter speed sufficient to avoid camera shake. We assume that the ambient illumination is weak, thus A will typically be very noisy and the illumination in F will be dominated by the dark flash $I^f(\lambda)$. We seek an image R whose edges are close to those in F and whose intensities are close to a denoised version of A, hopefully being similar to a long-exposure shot of the scene L.

Standard approaches to denoising use spatial priors that enforce sparsity on image gradients [Portilla et al. 2003]. In the flash/no-flash scenario, F contains high-frequency details that can assist the denoising process. But unlike conventional flash/no-flash photography, our flash and ambient illuminations $I^f(\lambda)$ and $I^a(\lambda)$ are by design almost non-overlapping, thus the colors in F will be quite different to those in the ambient image A or the long-exposure L. We propose a solution that uses the strong correlations between color channels as a constraint in an optimization scheme which computes R from A and F.

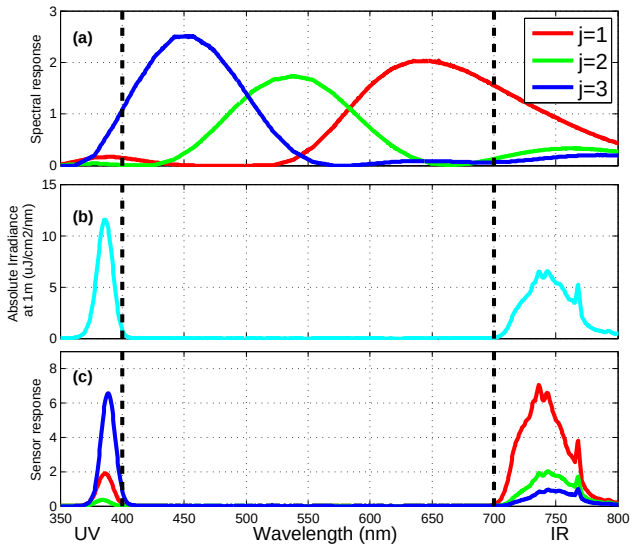


Figure 2: *(a) Spectral response curves $C_j(\lambda)$, $j = \{1,2,3\}$ for each of the camera's three color channels. (b) Absolute irradiance 1m from the dark flash $I^f(\lambda)$. (c) Spectrum received by the camera sensor when imaging a perfect white surface ($S(p,\lambda)$=1) illuminated by the dark flash. The curves are the product of those shown in (a) and (b). The recorded pixel values for the three channels are the integrals of these curves (see Eqn. 1). Note under the dark flash: no channel records in the visible range (black dashed lines); channel j=3 measures in the UV and channel j=1 responds to IR.*

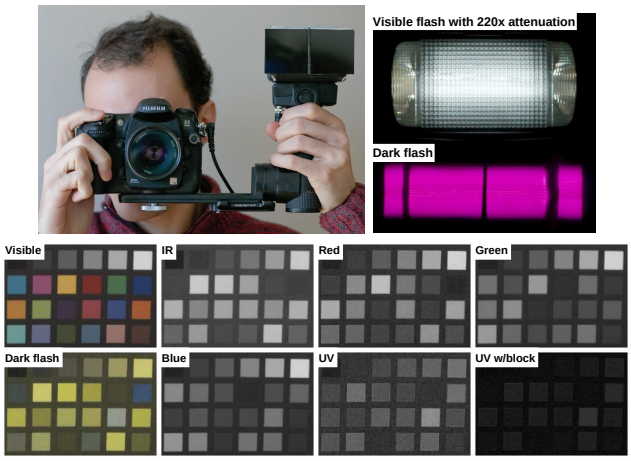


Figure 3: *Top left: Our camera and dark flash system. Top right: The perceived brightness of the dark flash and a visible flash that gives a comparable camera exposure. To capture them in a single image, it was necessary to attenuate the visible flash by a factor of 220 using neutral density filters. Without these, the dark flash would not be visible in a non-saturated 8-bit image. Bottom: A color chart captured with a pair of flash images (visible and dark), separated out into five spectral bands. The bottom right subplot shows the UV band with a UV-block filter attached to the camera that has a sharp cut-off at 400nm. The low intensities in this band show that our camera is genuinely recording UV light, not blue light from fluorescence caused by the UV part of the flash. See Section 4.2 for further discussion.*

3.1 Spectral constraints

Consider the 1-D example in Fig. 4 which shows a scanline across 3 squares in the color chart from Fig. 3. Fig. 4(a) shows the intensities from the red channel of a long exposure shot (L_1, in magenta) and IR from the dark flash (F_1, in black). Although the intensities are quite different, the edges are aligned, since the spectral reflectance at red and IR wavelengths are correlated with one another. The alignment of the edges is apparent in Fig. 4(b) where the gradients along the scanline ∇F_1 and ∇L_1 are shown ($\nabla F_1(p) = F_1(p) - F_1(p-1)$, the difference between adjacent pixels p). As is widely known, this gradient signal is sparse, being close to zero everywhere but a few locations. Now, if we consider the *difference* between the two gradient signals $\nabla F_1 - \nabla L_1$ (Fig. 4(c)) then this too will be sparse, as shown by shape of the histogram in Fig. 4(d). Now consider a dark flash and noisy ambient image pair, shown in Fig. 4(e)–(h). The difference between gradients $\nabla F_1 - \nabla A_1$ (in Fig. 4(g)) is now no longer sparse, as shown by it's Gaussian-shaped histogram in Fig. 4(h).

Reflecting the sparse distribution of $\nabla F_1 - \nabla L_1$ in Fig. 4(d), our spectral constraints take the form of a sparse norm on the gradient difference between channels in the reconstructed image R and the flash image F_1, i.e. $|\nabla R_j - \nabla F_1|^\alpha$ where $\alpha \leq 1$. This encourages the edge structures in R_j to align spatially with those in F_1 while allowing their magnitudes to differ. Thus, when transitioning between two materials, it does not matter if the spectral reflectances are different in visible and IR/UV bands, provided that there is an significant edge in IR/UV. If an ℓ_2 norm were used, this would not be the case, and ∇R_j and ∇F_1 would have to closely match, even at material transitions, so causing artifacts in R_j (see Fig. 9). While a conventional spatial prior, such as $|\nabla R_j|^\alpha, \alpha < 1$, would also reduce noise, it would not encourage the edges to align with those of F which are close to those of the desired solution L.

We also impose a similar constraint to the UV channel: $|\nabla R_j - \nabla F_3|^\alpha$, recalling that F_3 records UV and F_1 records IR. For R_3 (the blue channel), this will be a strong constraint since, in terms of wavelength, blue is much closer to UV than to IR. In this example, we have only considered 1-D gradients but in the real problem we use both x and y gradients, with separate terms for each. For brevity, we use ∇ to refer to both ∇_x and ∇_y.

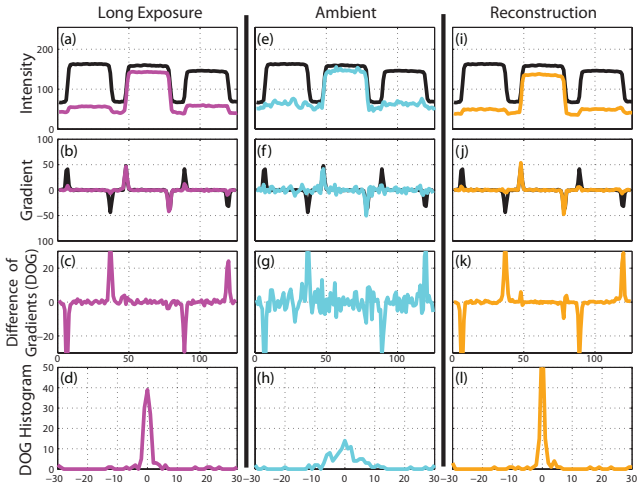


Figure 4: *1-D example of the spectral constraints in our model, using a a scan line across 3 squares in the color chart of Fig. 3. See text for explanation.*

3.2 Spatial-spectral cost function

Our cost function consists of three main terms: (i) **Likelihood**: the intensities of the reconstruction R_j should be close to those of the noisy ambient image A under an ℓ_2 norm, assuming a Gaussian noise model. (ii) **Spatial prior**: ∇R_j should be small under a sparse norm, reflecting the heavy-tailed nature of image gradients. The spatial prior term helps to give a further boost to image quality. (iii) **Spectral constraint**: ∇R_j should be close to both ∇F_1 (IR) and ∇F_3 (UV) under a sparse norm, as explained above.

As with existing flash/no-flash techniques, we use a shadow and specularity mask $m(p)$ which removes artifacts from the flash image. Details of the mask construction are given in Section 3.3 below. The overall cost function for each channel j is:

$$\underset{R_j}{\operatorname{argmin}} \sum_p \Big[\underbrace{\mu_j\, m(p)(R_j(p) - A_j(p))^2}_{\text{Likelihood}} + \underbrace{\kappa\, m(p)|\nabla R_j(p)|^\alpha}_{\text{Spatial}} + \underbrace{|\nabla R_j(p) - \nabla F_1(p)|^\alpha}_{\text{IR Spectral}} + \underbrace{|\nabla R_j(p) - \nabla F_3(p)|^\alpha}_{\text{UV Spectral}} \Big] \quad (2)$$

In our experiments, unless otherwise stated, we use $\kappa = 1$, $\alpha = 0.7$. We solve for each each channel j separately. $m(p)$ has the effect of increasing the weight on the likelihood and spatial terms in regions of shadows or specularities. We also assumed the UV and IR spectral terms to have equal weight for all channels j. Hence the weighting on the reconstruction term for each channel μ_j is the only important parameter in the model and strongly depends on the noise level of the ambient image A. Since the blue channel is often significantly noisier than the others, we use a different value for μ_3 than for μ_1 and μ_2 (which are set to be the same). Intuitively, if μ_j is set to a large value then the colors of R will be close to those of A at the expense of increased noise. Conversely, if μ_j is small then the noise in R is reduced, but the colors will deviate from those in A. Choosing the value of μ_j can be done semi-automatically from

the level of under-exposure of A (given by the camera's exposure meter) and the camera's ISO setting. If needed, the value may be fine-tuned on a small image patch, before processing the entire image. Typical values range from $\mu_j = 5$ (high noise) to $\mu_j = 40$ (low noise).

Returning to our 1-D example in Fig. 4, we show the scanline across the color chart for our reconstructed image R in Fig. 4(i)–(l). Despite the spectral reflectances of the squares being quite different, the intensities of R_1 shown in orange in Fig. 4(i) closely match those of the desired solution L_1 in Fig. 4(a). Note that R_1 is kept close to A_1 (shown in Fig. 4(e)) by the likelihood term, while the sparse norm on the spectral terms removes the noise.

We optimize Eqn. 2 (which is non-convex if $\alpha = 0.7$) using Iterative Re-weighted Least Squares [Levin et al. 2007], initializing with $R_j = F_j$. Due to poor conditioning of the least-squares systems, we use an incomplete Cholesky preconditioner to speed convergence. For a 1.3 megapixel image, our unoptimized Matlab implementation takes approximately 25 minutes for all 3 channels, with 5 iterations/channel. As this may be unacceptably slow for some practical situations, a considerable speedup can be achieved by setting $\alpha = 1$. This makes the problem convex and fast numerical schemes can be used (e.g. [Wang et al. 2008b]), resulting in a processing time of 3 minutes, comparable to efficient implementations of the cross-bilateral filter. However, some image quality is lost in using $\alpha = 1$ and we explore this issue further in Fig. 9.

3.3 Pre & post-processing

Pre-processing. All images were captured in RAW mode. They were then demosaiced and manually white-balanced using some neutral-colored object (e.g. a wall or calibration target) in the scene. The mask $m(p)$ was built using the same methods used in [Petschnigg et al. 2004], namely the shadows were detected by finding areas where $|F - A|$ is very small. Specularities were found by looking for pixels saturated in F_1 (IR channel). In areas of shadow/specularity $m(p) = 5$ and $m(p) = 1$ in all other areas, smoothly varying between the two at the boundaries. In high noise conditions, we apply a small Gaussian smoothing to A_j to break up any spurious image structure formed by the noise. The optimization is then performed on the linear tonescale images (i.e. without gamma correction).

Post-processing. If the ambient light levels are very low, the colors in the ambient image can become imbalanced, particularly with a blue tint due to excessive noise levels in the blue channel. Hence the output of the optimization will also have a similar color cast and will not look similar to a long-exposure shot L. To compensate for this, we use an additional color correction operation that applies a global color mapping to R. To generate this mapping function, we determined the tone response curve of our camera for each color channel using a stack of images taken over a wide range of exposures [Debevec and Malik 1997]. Particular care was taken when fitting the parametric model to the low intensity part of the curve. In this regime, the sensor noise causes the curve to be non-linear, in turn giving rise to the color casts observed in very noisy images (e.g. Fig. 5). By passing each R_j through its appropriate mapping function, we can infer the true value of each pixel, yielding colors close to those in a long-exposure shot L. Finally, we gamma-correct the images for display, using $\gamma = 1.8$.

4 Results

For the dark flash system to be practical it must achieve high quality reconstructions in low levels of ambient illumination. In Fig. 5 and Fig. 6 we show 4 test examples: two portrait shots and two still scenes. The test images were captured using two different types of ambient illumination (tungsten and compact fluorescent) and contain a wide range of materials and colors. The images in Fig. 5 and Fig. 6 are high resolution so are best viewed under magnification, in order that fine details and noise may be seen. To show how the noise levels vary across color channel we show a small region in two of the images, separated out into its constituent color planes. This typically reveals the blue channel to be far noisier than the others.

To make comparisons straightforward, the shutter speed used to capture the flash/no-flash pair is varied, thus simulating different levels of ambient illumination. In practice however, the shutter speed would be set to the slowest level that avoids camera shake, irrespective of the level of ambient light. As the light levels drop, the ambient image becomes noisier (the dark flash image F stays constant, however) thus making the reconstruction harder. Three different noise scenarios are explored: (i) Low, where it is possible to achieve reconstructions close to a long exposure reference shot; (ii) Medium, where the reconstruction is acceptable in terms of quality and (iii) High, where a significant degradation in quality is visible and the failure modes of the algorithm are evident. At each noise level, the degree of under-exposure of the ambient image A, relative to the long exposure reference L, is quoted. These range from 1/32nd of ambient illumination (Fig. 6(top)), down to 1/256th for the portrait shots. Assuming 1/30th of a second is required to avoid camera shake, the results are equivalent to taking pictures in conditions where exposures ranging from 1 second to 8 seconds would otherwise be required. Techniques that permit blur-free photography at slow shutter speeds, such as image stabilizers, would extend the range of operation of the dark flash system to even longer equivalent exposures.

Ensuring accurate alignment between F and A is an important practical issue since the spectral constraints require this. While a range of software approaches for image registration exist (e.g. [Baker et al. 2004]), any commercial implementation of the system would use a hardware approach based on sensors that can capture pairs of images with virtually no delay between them (e.g. Fuji Finepix Z10fd), guaranteeing good alignment. Thus with our prototype, we sidestep this issue and capture the shots using a tripod. It is difficult to draw comparisons with Petschnigg et al.[2004] since they do not specify the exposures used to capture their images, but qualitatively the majority of their examples correspond to our low noise case, with a single case being equivalent to our medium noise level.

At high noise levels, some color deviations and loss of detail can be observed. This is a consequence of low μ_j values which give the likelihood term little weight in the optimization. At all noise levels, our reconstructions contain some artifacts that result from the dark flash illumination. If a material absorbs both UV and IR strongly, then F will contain no gradients to guide the reconstruction. Examples of this include: the freckles on the man in Fig. 1 & Fig. 5(lower) and the red lips of the doll in Fig. 6. Fortunately, this is relatively uncommon, as demonstrated by the range of colors and materials in our shots, the vast majority of which are accurately recovered. In particular, human skin and hair, two materials relevant to the dark flash application, are plausibly reproduced.

4.1 Comparison experiments

We compare our method to a range of different hardware and software approaches. In Fig. 7 we explore in turn the importance of having UV and IR in our dark flash by removing the corresponding spectral term in the cost function of Eqn. 2. The figure shows the need for both the UV and IR components, since if either is removed, the adjacent spectral bands (blue and red, respectively) in R become degraded.

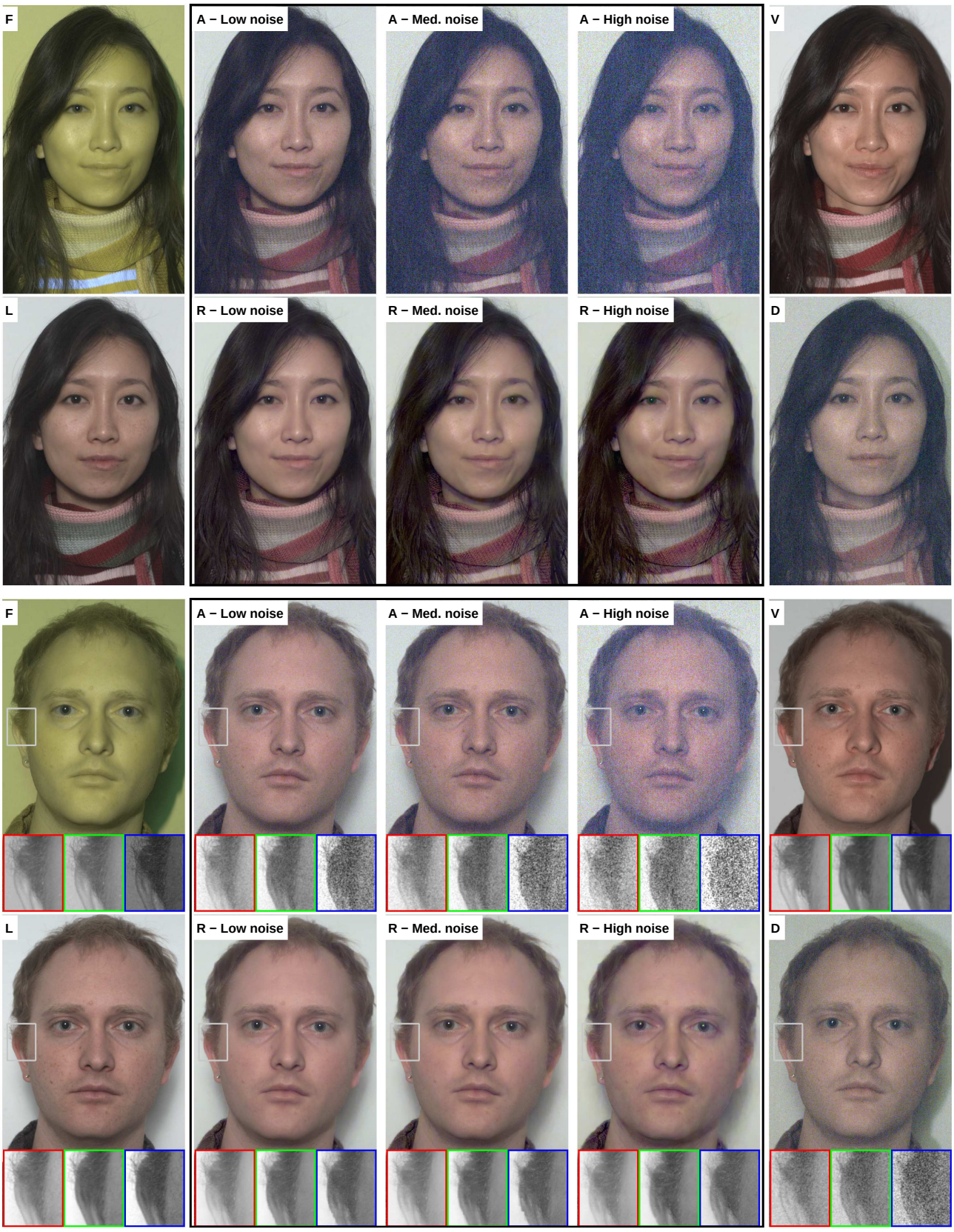


Figure 5: *Two portrait shots captured with our camera/flash under tungsten illumination. Within each group, column 1 shows the dark flash shot (F) and long exposure reference (L). Our results are shown in Columns 2,3 & 4. For each ambient image (A) of decreasing exposure (yielding increased noise), we show the reconstructed output (R). Column 5 shows a visible flash image (V), along with a visible flash shot (D) attenuated with neutral density filters so that it is comparably dazzling to F. The Low, Medium and High noise levels correspond to 6, 7 and 8 stops of underexposure respectively (corresponding to 1/64th, 1/128th and 1/256th of ambient long exposure). In the lower group, we show a zoomed-in section, separated into red, green, blue color channels.*

Figure 6: *Two different scenes captured with our camera/flash under fluorescent illumination. Within each group, rows 1 & 2 show shots under ambient illumination (A) of decreasing exposure (yielding increased noise) and our reconstructed output (R). Row 3 shows, from left to right: Long exposure reference (L), Visible flash shot (V) and dark flash shot (F). In the top group, Low, Medium and High noise levels correspond to 5, 6 and 7 stops of underexposure respectively (equating to 1/32nd, 1/64th and 1/128th of ambient long exposure). In the bottom, Low = 5.5, Medium = 6.5 and High = 7.5 stops underexposed (corresponding to 1/45th, 1/90th and 1/180th of ambient).*

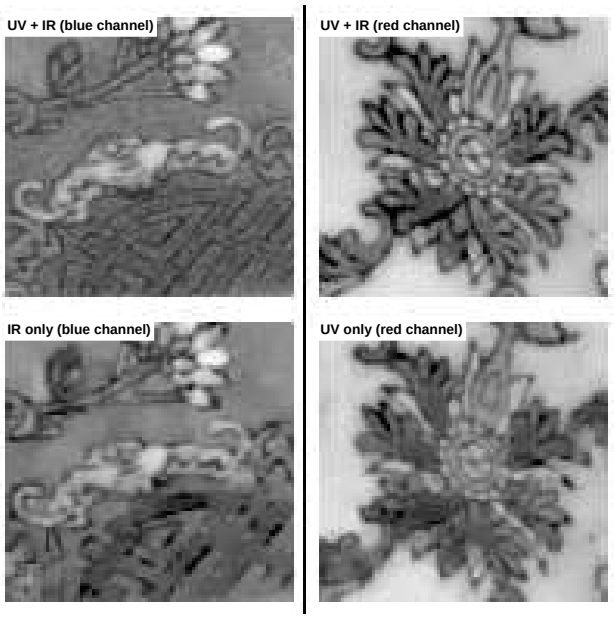


Figure 7: *Closeup of Fig. 6 (bottom group), showing the need for both spectral terms in Eqn. 2. Top left: Blue channel of reconstructed image R using both UV and IR spectral terms. Bottom left: Blue channel using only IR spectral term. Top right: Red channel of reconstructed image R using both UV and IR spectral terms. Bottom right: Red channel using only UV spectral term. Note that the removal of the flash in the adjacent band causes a degraded result.*

In Fig. 8 we compare our algorithm to alternate methods, using the mid-noise case. First, we use the processing pipeline based on the cross-bilateral filter and detail enhancement, as described in [Petschnigg et al. 2004]. Using the dark flash/ambient image pair with their system, the results obtained are inferior to our approach. The range term in the cross-bilateral filter causes the edge strength in the flash image F to directly influence the smoothing of the ambient image A. Thus it will only operate correctly if the edges in F and A are closely matched in magnitude, an unrealistic assumption since spectral reflectances typically differ between bands. In contrast, our model permits the edge magnitudes to differ when $\alpha \leq 1$ in Eqn. 2, giving a reconstruction of superior quality. Second, we tried two approaches that attempt to directly denoise the ambient image: (i) bilateral filtering [Tomasi and Manduchi 1998] and (ii) a commercial denoising tool, Noise Ninja [Christian and Zapata 2008]. Both methods perform poorly compared to the flash/no-flash approaches.

In Fig. 9 we explore how the value of α in Eqn. 2 effects the reconstruction. When a non-sparse norm is used ($\alpha = 2$), the ambient colors bleed. This can be prevented by using $\alpha \leq 1$, with some improvement in quality for $\alpha = 0.7$.

4.2 Fluorescence

Certain materials fluoresce when illuminated by the UV component of our flash, the most common instances being white items of clothing such as the stripes in Fig. 5(top). Fluorescence manifests itself as visible blue light that gives an unnaturally bright intensity in F_3 in that part of the scene. Experimentally, we find the phenomenon to be relatively rare: our test scenes contain a wide range of materials, natural and man-made, yet it only occurs in a few locations. It is certainly not the dominant source of signal in F_3, as demonstrated by Fig. 3(bottom). Where it does occur, it can produce some minor

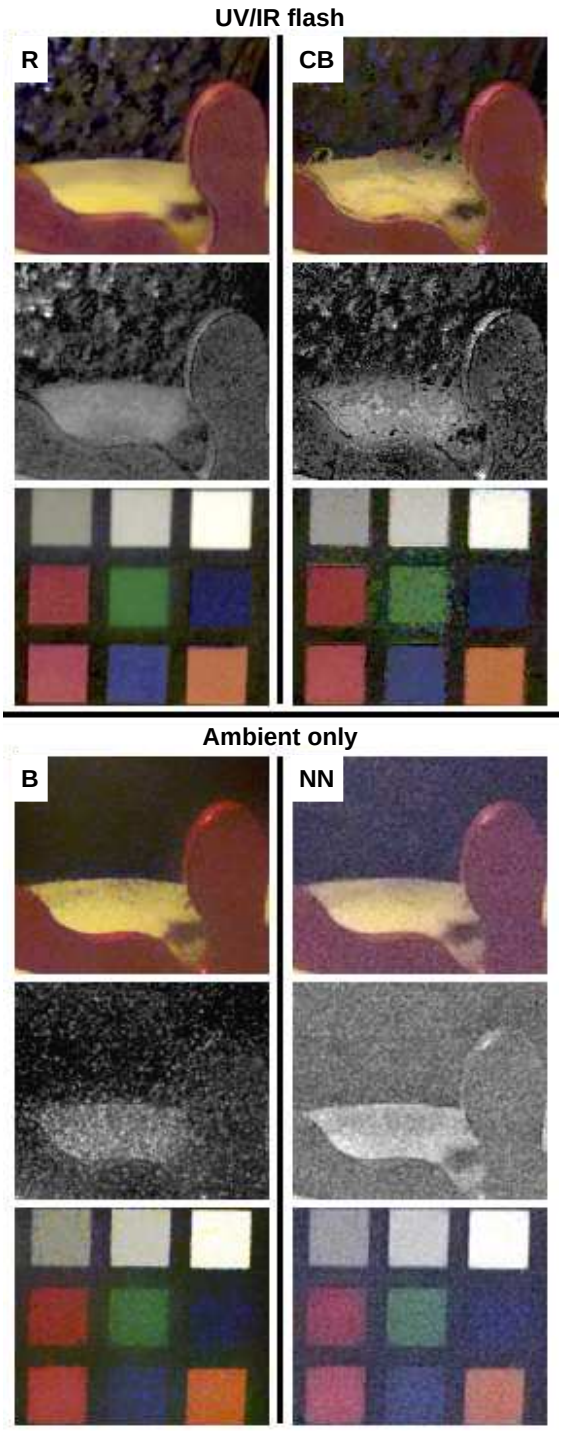


Figure 8: *Comparison of our approach to different processing methods, showing two crops from Fig. 6 (top group), along with the blue channel of the first crop. The top set uses a dark flash / ambient image pair, while the bottom uses the ambient image only. Key. R: Our reconstruction using spectral constraints. CB: Pipeline from [Petschnigg et al. 2004] based on cross-bilateral filter and detail enhancement. B: Bilateral filter of ambient image [Tomasi and Manduchi 1998]. NN: Noise Ninja commercial denoising plugin for Photoshop [Christian and Zapata 2008]. Our reconstruction approach produces superior results to the cross-bilateral approach and the standard denoising methods.*

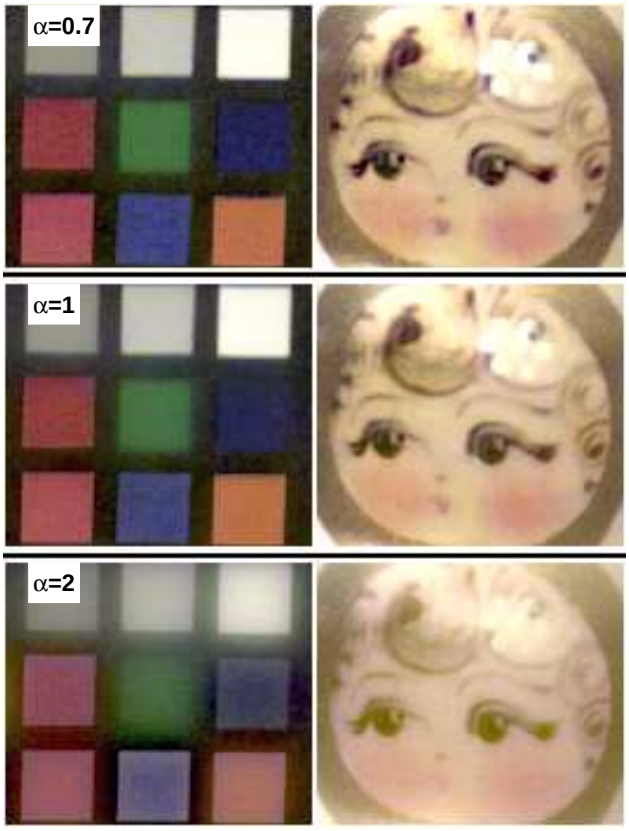


Figure 9: *Effect of varying α in Eqn. 2. For values ≤ 1, R contains crisp edges, even if the spectral reflectances of the materials in visible and non-visible wavelengths differ somewhat, as is typically the case. Setting $\alpha = 2$ has the undesirable effect of causing the colors to bleed between regions. When $\alpha = 2$ the spectral constraints force the edges in the UV/IR flash and ambient to be the same, an unrealistic assumption given that they are captured at different wavelengths.*

purple artifacts. Another drawback is that other people observing the subjects during the photograph may see a glow from the clothing, thus making the flash not so invisible to them, although the subjects themselves, if looking at the camera, will not notice this.

4.3 Photometric flash measurements

One of the main objectives of our dark flash is that it should be as unnoticeable as possible to human subjects. We measured the dark flash output with a spectrometer to determine the spectral irradiance (shown in Fig. 2(b)) 1m from the flash. This was then converted to photometric units, using the photopic luminosity function of Vos [1978]. The luminous exposure for the dark flash was 1.6 lux seconds. A visible flash set to produce an image V of similar intensity to a dark flash image F had luminous exposure of 362 lux seconds, a factor of 226 times brighter. This ratio agrees closely with the experiment of Fig. 3(top right) where an attenuation of 220 times was required to make the visible flash of comparable brightness to the dark flash. In Fig. 5, we show images D captured with a visible flash attenuated by this factor. The resulting images are unacceptably noisy.

Subjectively, people report that when looking directly at the flash they see a weak purple light that does not dazzle, or leave an afterimage. They also report that if not looking directly at the dark flash, the burst of light is very easy to miss. By contrast, when using a visible flash that gives a comparable scene exposure, the burst of light is highly dazzling and leaves a strong after-image.

5 Other applications

Although our main focus has been the dark flash application, both the hardware and software elements of our system can be used in a variety of other ways.

5.1 Estimation of spectral reflectance

By taking two images, one with the dark flash, the other with a visible flash, we can obtain 5 different spectral measurements at each point in the scene: UV,B,G,R,IR as opposed to 3 obtained with a conventional camera. The spectral reflectances of real world materials can be accurately modeled in a low-dimensional subspace using PCA with relatively few components [Wandell 1995]. Using a spectrometer and reflectance probe, we measured 255 different materials in the real world and computed a set of 5 PCA basis functions for the range 360–800nm. We then used the constrained least squares formulation introduced in [Park et al. 2007] to solve for the spectral reflectance functions for all points in the scene ($S(p, \lambda)$ in Eqn. 1). In Fig. 10(left), we show the estimated spectral reflectance for four squares from the color chart in Fig. 6 (top group), along with ground truth. Note that we are able to accurately infer the spectrum beyond the visible range. In Fig. 10(right) we compare the RMS error between our spectra and the ground truth over the visible range. We achieve very similar total error to the approach of Park et al.[2007]: 0.82 and 0.79 respectively, compared to 1.19 when using R,G,B channels alone.

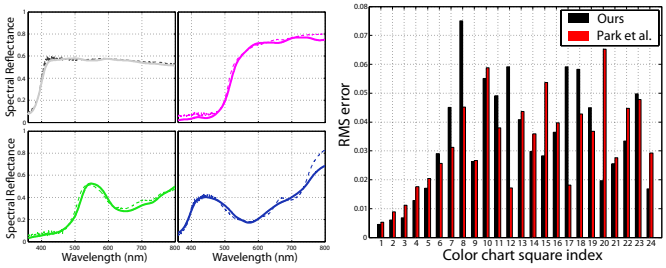


Figure 10: *Using a dark/visible flash pair we are able to accurately infer the spectral reflectance of objects. Left: Spectra of four different squares from the color chart in Fig. 6. Solid line is inferred spectrum, dashed line is ground truth. Line colors correspond to square color. Right: RMS estimation errors for all 24 squares in color chart over 400-700nm range, compared to results of multi-spectral illumination approach of Park et al.[2007].*

5.2 Color-band denoising

The spectral constraints used in our dark flash approach can be applied to images captured by standard cameras. One example, as shown in Fig. 11, is for conventional flash/no-flash processing, using a visible flash/ambient pair. When using our algorithm in this configuration, the spectral constraint reduces to a single term linking each channel in the flash image to its corresponding channel in the ambient, hence the term no longer links between different spectral bands. Our algorithm yields better results than the cross-bilateral based method.

Another application is where one color channel is much noisier than the others. For example, candle-light is very weak in the blue part of the spectrum, compared to red and green. Hence when trying

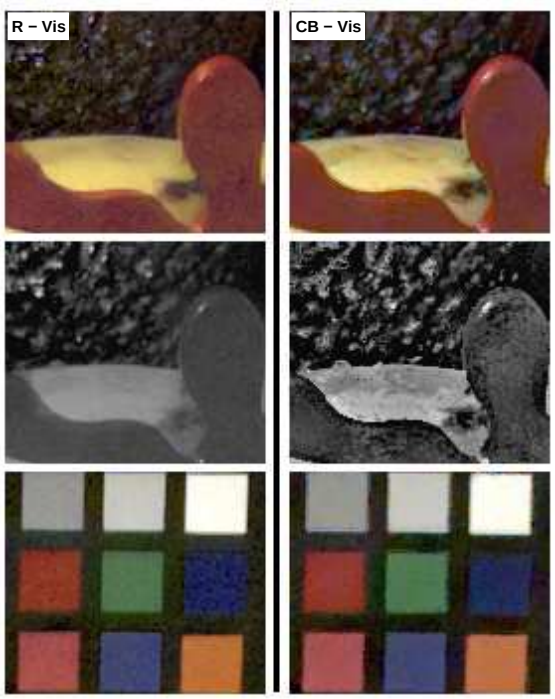


Figure 11: *The model in Eqn. 2 being used in a visible flash/no-flash setting. The two crops are taken from Fig. 6 (top group), with the center row showing the the blue channel of the first row.*
R - Vis: reconstruction with our model using spectral constraints.
CB - Vis: Pipeline from [Petschnigg et al. 2004] based on cross-bilateral filter and detail enhancement.

to white balance a candle-lit image, the blue channel must be multiplied by a large factor, increasing the noise levels. Using spectral constraints, the blue channel can be denoised using the red and green channels (in place of F_1 and F_3 in Eqn. 2). This gives a superior result to denoising the blue channel using spatial priors and likelihood alone. See Fig. 12 for this technique applied to a candle-lit image captured with an unmodified Canon 40D.

6 Discussion

We have demonstrated a camera and flash system that can take pictures in low light conditions using a flash that is far less noticeable and disruptive than a conventional one. The system uses standard hardware for the most part, combined with novel image processing techniques. The spectral constraints are a powerful way of combining the images, yielding good quality results in low light conditions. In addition, we have shown that the hardware and software techniques introduced in this paper can be used in a number of other applications.

Our hardware is a prototype and can be improved in a number of ways. An obvious limitation is the need to take two images of the scene. This precludes the capture of fast moving scenes and adds to the overall complexity of the system. However, by modifying the Bayer pattern on the sensor to include UV-only and IR-only pixels (for a total of 5 channels), we would be able to implement the dark flash concept using a single image. Additionally, our large flash unit could be replaced with compact UV and IR LEDs giving a more controllable pulse duration and a more precise spectral emission, perhaps further reducing the visibility of the flash. This would also

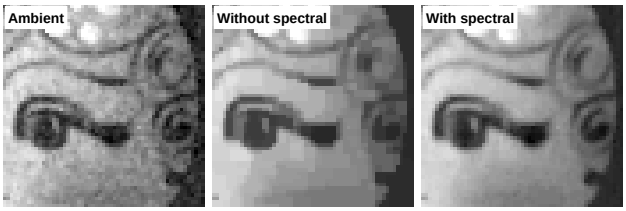


Figure 12: *Close up of scene in Fig. 6 (top group) illuminated by candlelight. Left: Blue channel of white-balanced ambient shot, showing high noise due to lack of blue wavelengths in candle-light. Middle: Denoising of ambient using likelihood and spatial priors only. Right: Denoising of ambient using spectral constraints from the red and green channels, in addition to the likelihood and spatial priors. The spectral constraints significantly improve performance.*

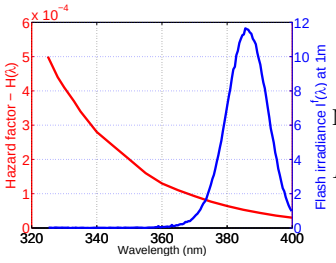
permit the dark flash concept to be implemented in small platforms such as cell-phones, where a flash is often needed due to poor low-light performance on account of the small sensor size.

Appendix A

We now give hardware details of our camera and flash system. All experiments used a standard Nikon 50mm f/1.8 lens, which transmits light down to 350nm, hence is not the limiting factor in the camera's UV response. A MaxMax CC3 filter was attached to the lens at all times. The purpose of this filter is to block IR light above 850nm, which would otherwise distort the colors of the ambient image (as the naked sensor's response extends out to 1100nm). This filter does not block either visible light or the dark flash. The response functions $C_j(\lambda)$ in Fig. 2(a) include the filter and lens. The flash is a clone of the Nikon SB-14UV, adapted from a standard SB-14 by removing the UV absorbent coating on the Xenon flash tube. A Hoya U360 filter was attached to the flash at all times to filter out visible light. The standard visible flash used in comparisons was equipped with a MaxMax CC1 filter to block its significant IR output.

Appendix B

We now detail the safety calculations summarized in Section 2. The threshold limit values (TLVs) for UV radiation 180–400nm incident on the eye (the most sensitive part of the body) over any 8 hour period are given by the formula on p.155 of [TLVs 2001], reproduced in Eqn. 3 below. It relates the maximum number of flashes to the effective irradiance E_{Eff}, relative to a monochromatic source at 270nm. E_{Eff} is computed, using Eqn. 4 below, from the spectral irradiance of the flash $I^f(\lambda)$ (units: μJ/cm^2/nm/flash) and a hazard weighting function $H(\lambda)$ (which is 1 at 270nm), given on p.157 of [TLVs 2001]. In Fig. 13, we show $I^f(\lambda)$ and $H(\lambda)$. Integrating over the product of the two and inserting E_{Eff} into Eqn. 3, we arrive at the value of 130,000 flashes. Note that this number scales with the inverse square of distance, so at 2m the max safe limit would be 520,000 flashes.



$$\text{Max flashes} = \frac{3000}{E_{\text{Eff}}} \quad (3)$$

$$E_{\text{Eff}} = \int H(\lambda) I^f(\lambda)\, d\lambda \quad (4)$$

Figure 13: $I^f(\lambda)$ *and* $H(\lambda)$*, see text for details.*

Acknowledgments

This work was supported by a gift from Microsoft Research. The authors are indebted to the anonymous reviewers, as well as Fredo Durand, Anat Levin, Yann LeCun, Olga Sorkine and Denis Zorin for their insights and suggestions. We would also like to thank Murphy Stein and Agnes Chen for posing as subjects and Dan Llewellyn of LDP LLC (MaxMax.com) for hardware assistance.

References

AGRAWAL, A., RASKAR, R., NAYAR, S., AND LI, Y. 2005. Removing photography artifacts using gradient projection and flash-exposure sampling. In *ACM Transactions on Graphics (Proc. SIGGRAPH)*, vol. 24, 828–835.

AHARON, M., ELAD, M., AND BRUCKSTEIN, A. 2006. The K-SVD: An algorithm for designing of overcomplete dictionaries for sparse representation. *IEEE Trans. Signal Processing 54*, 11 (November), 4311–4322.

BAKER, S., GROSS, R., AND MATTHEWS, I. 2004. Lucas-kanade 20 years on: A unifying framework. *International Journal of Computer Vision 56*, 221–255.

BENNETT, E., MASON, J., AND MCMILLAN, L. 2007. Multispectral bilateral video fusion. *IEEE Trans. Image Processing 16*, 5, 1185–1194.

CHAKRABARTI, A., HIRAKAWA, K., AND ZICKLER, T. 2008. Color constancy beyond bags of pixels. In *CVPR*, 1–6.

CHRISTIAN, J., AND ZAPATA, F., 2008. Noise Ninja, Photoshop denoising plugin. `http://www.picturecode.com/`.

DEBEVEC, P. E., AND MALIK, J. 1997. Recovering high dynamic range radiance maps from photographs. *ACM Transactions on Graphics (Proc. SIGGRAPH) 31*, 3, 369–378.

EISEMANN, E., AND DURAND, F. 2004. Flash photography enhancement via intrinsic relighting. In *ACM Transactions on Graphics (Proc. SIGGRAPH)*, vol. 23, 673–678.

FARBMAN, Z., FATTAL, R., LISCHINSKI, D., AND SZELISKI, R. 2008. Edge-preserving decompositions for multi-scale tone and detail manipulation. In *ACM Transactions on Graphics (Proc. SIGGRAPH)*, vol. 27, 671–680.

FERGUS, R., SINGH, B., HERTZMANN, A., ROWEIS, S. T., AND FREEMAN, W. 2006. Removing camera shake from a single photograph. *ACM Transactions on Graphics (Proc. SIGGRAPH) 25*, 787–794.

JIAYA, J. 2007. Single image motion deblurring using transparency. In *CVPR*, 1–8.

LEVIN, A., AND WEISS, Y. 2007. User assisted separation of reflections from a single image using a sparsity prior. *IEEE Trans. Pattern Analysis and Machine Intelligence 29*, 9 (Sept), 1647–1654.

LEVIN, A., FERGUS, R., DURAND, F., AND FREEMAN, W. 2007. Image and depth from a conventional camera with a coded aperture. *ACM Transactions on Graphics (Proc. SIGGRAPH) 26*, 3, 70.

MCAULEY, J. J., CAETANO, T. S., SMOLA, A. J., AND FRANZ, M. O. 2006. Learning high-order MRF priors of color images. In *ICML '06*, 617–624.

MOHAN, A., RASKAR, R., AND TUMBLIN, J. 2008. Agile spectrum imaging: Programmable wavelength modulation for cameras and projectors. *Computer Graphics Forum 27*, 2, 709–717.

MORRIS, N., AVIDAN, S., MATUSIK, W., AND PFISTER, H. 2007. Statistics of infrared images. In *CVPR*, 1–7.

PARK, J., LEE, M., GROSSBERG, M. D., AND NAYAR, S. K. 2007. Multispectral Imaging Using Multiplexed Illumination. In *ICCV*, 1–8.

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

PORTILLA, J., STRELA, V., WAINWRIGHT, M. J., AND SIMONCELLI, E. P. 2003. Image denoising using a scale mixture of Gaussians in the wavelet domain. *IEEE Trans. Image Processing 12*, 11 (November), 1338–1351.

RORSLETT, B., 2008. `http://www.naturfotograf.com/UV\_flowers_list.html`.

ROTH, S., AND BLACK, M. J. 2005. Fields of Experts: A Framework for Learning Image Priors. In *CVPR*, vol. 2, 860–867.

SINGH, B., FREEMAN, W. T., AND BRAINARD, D. H. 2003. Exploiting spatial and spectral image regularities for color constancy. In *Workshop on Statistical and Computational Theories of Vision*.

TELLEEN, J., SULLIVAN, A., YEE, J., WANG, O., GUNAWARDANE, P., COLLINS, I., AND DAVIS, J. 2007. Synthetic shutter speed imaging. *Computer Graphics Forum 26*, 3 (Sept.), 591–598.

TLVS. 2001. *TLVs and BEIs: threshold limit values for chemical substances and physical agents*. American Conference of Governmental Industrial Hygienists.

TOMASI, C., AND MANDUCHI, R. 1998. Bilateral filtering for gray and color images. In *ICCV*, 839–846.

VOS, J. 1978. Colorimetric and photometric properties of a 2-deg fundamental observer. *Color Research and Application*, 125–128.

WANDELL, B. A. 1995. *Foundations of Vision*. Sinauer Associates.

WANG, O., DAVIS, J., CHUANG, E., RICKARD, I. AMD DE MESA, K., AND CHIRAG, D. 2008. Video relighting using infrared illumination. *Computer Graphics Forum 27*.

WANG, Y., YANG, J., YIN, W., AND ZHANG, Y. 2008. A new alternating minimization algorithm for total variation image reconstruction. *SIAM J. Imaging Sciences 1*, 3, 248–272.

YUAN, L., SUN, J., QUAN, L., AND SHUM, H.-Y. 2007. Image deblurring with blurred/noisy image pairs. In *ACM Transactions on Graphics (Proc. SIGGRAPH)*, vol. 26, 1–10.