

# Supplemental Materials: Efficient Multispectral Facial Capture With Monochrome Cameras

Chloe LeGendre  
USC Institute for Creative  
Technologies

Xinglei Ren  
USC Institute for Creative  
Technologies

Kalle Bladin  
USC Institute for Creative  
Technologies

Xueming Yu  
Google

Bipin Kishore  
USC Institute for Creative  
Technologies

Paul Debevec  
USC Institute for Creative  
Technologies

## ACM Reference format:

Chloe LeGendre, Kalle Bladin, Bipin Kishore, Xinglei Ren, Xueming Yu, and Paul Debevec. 2018. Supplemental Materials: Efficient Multispectral Facial Capture With Monochrome Cameras. In *Proceedings of SIGGRAPH '18 Posters, Vancouver, BC, Canada, August 12-16, 2018*, 6 pages. <https://doi.org/10.1145/3230744.3230778>

## 1 METHOD: MULTISPECTRAL POLARIZATION PROMOTION

We develop a technique that we call *multispectral polarization promotion* in which we *hallucinate* cross-polarized images for each spectral channel from unpolarized lighting images, so that we can generate a multispectral diffuse albedo texture map of the subject. Our process requires that only one of the spectral channels in the lighting rig is polarized in the pattern of Ghosh et al. [2011].

For clarity, we extend the variable naming conventions of Ma et al. [2007]. We define a gradient illumination image of the subject  $L_{l,i,s}$ , where  $l$  describes the gradient condition,  $i$  describes the polarization state (one of cross or parallel), and  $s$  defines the index of spectrum of illumination, ranging from 0 to  $n - 1$  where  $n$  is the number of spectral channels in the lighting rig, and 0 represents the white LED. The gradient illumination images required [Ma et al. 2007] are therefore:

- $L_{x,c,0}$ , cross-polarized, x gradient
- $L_{y,c,0}$ , cross-polarized, y gradient
- $L_{z,c,0}$ , cross-polarized, z gradient
- $L_{f,c,0}$ , cross-polarized, full sphere
- $L_{x,p,0}$ , parallel-polarized, x gradient
- $L_{y,p,0}$ , parallel-polarized, y gradient
- $L_{z,p,0}$ , parallel-polarized, z gradient
- $L_{f,p,0}$ , parallel-polarized, full sphere

When linear polarizers over the light sources are oriented perpendicularly to the those in front of the camera, the polarizer will block all of the specularly-reflected light and about half of the diffusely reflected light, such that  $L_{l,i=c,s} = \frac{1}{2}D_{l,s}$ , representing an image of the diffuse or sub-surface scattered reflections. When

the polarizer in front of the camera is parallel, the polarizer will block about half of the diffusely reflected light, and none of the specularly-reflected, such that  $L_{l,i=p,s} = \frac{1}{2}D_{l,s} + S_{l,s}$ . Therefore, for each gradient lighting condition  $l$  and spectrum  $s$ , the specular reflection image  $S_{l,s}$  is produced via polarization differencing:

$$S_{l,s} = L_{l,p,s} - L_{l,c,s} \quad (1)$$

Using a monochrome spectral camera model, a pixel value  $p_{s,j}$  of a material  $j$  lit by spectrum  $s$  is produced by integrating a fully-spectral modulation of the scene illuminant  $I_s(\lambda)$  by the reflectance spectrum of the material  $R_j(\lambda)$  and the monochrome camera's spectral sensitivity function  $C(\lambda)$ :

$$p_{s,j} = \int_{400}^{700} I_s(\lambda)R_j(\lambda)C(\lambda) \quad (2)$$

We again assume that light reflected specularly from the skin preserves both the polarization and spectrum of the incident source. This assumption implies for an image pixel representing specular reflection that the reflectance spectrum  $R_j(\lambda)$  of Eq. 2 is a constant value over the visible wavelength range. This value represents the per-pixel reflectivity or specular albedo ( $\rho_{spec}$ ) of the surface, modulated by a per-pixel constant scale factor  $F_l$  that only depends on the geometry of the illumination relative to the geometry of the surface. The intuition behind the constant  $F_l$  is that a different amount of light will be reflected specularly towards the camera for a pixel depending on the incident illumination condition  $l$  and the pixel's surface normal. Both constants can be pulled out from the integral, and the pixel values of the specular reflection image  $S_{l,s}$  are computed as:

$$S_{l,s} = (\rho_{spec}F_l) \int_{400}^{700} I_s(\lambda)C(\lambda) \quad (3)$$

In Eq. 3, the integral represents the intensity of  $I_s(\lambda)$  as observed by the monochrome camera with spectral response  $C(\lambda)$ . We call this quantity  $W_s$ :

$$W_s = \int_{400}^{700} I_s(\lambda)C(\lambda) \quad (4)$$

$W_s$  can be directly measured as a calibration step by photographing a reflective white spectralon disk or the white square of a color chart as lit by each spectrum of illumination  $s$  (scaled up to represent the true reflectance of these calibration targets). No spectral measurements are required. By substitution, we can write that the specular reflection image  $S_{l,s}$  is a scaled multiple of the incident light intensity, depending on the per-pixel specular albedo and

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

SIGGRAPH '18 Posters, August 12-16, 2018, Vancouver, BC, Canada

© 2018 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-5817-0/18/08.

<https://doi.org/10.1145/3230744.3230778>

per-pixel geometric factor:  $S_{l,s} = (\rho_{spec} F_l) W_s$ . Or, by rearranging:

$$(\rho_{spec} F_l) = \frac{S_{l,s}}{W_s} \quad (5)$$

We can equate these ratios across spectral channels for a given gradient illumination condition  $l$ . Without loss of generality, we can compare the white LED with another spectrum  $s$ :

$$\frac{S_{l,0}}{W_0} = \frac{S_{l,s}}{W_s} \quad (6)$$

We assume that with our lighting rig we are able to capture cross- and parallel-polarized images  $L_{f,c,0}$  and  $L_{f,p,0}$  for the white LED for the full sphere lighting condition  $f$ , producing  $S_{f,0}$  using Eq. 1. After measuring  $W_s$  for each spectral channel, we therefore solve for  $S_{f,s}$  for each spectral channel, by substitution into Eq. 6. The intuition behind this step is again that the amount of light reflected specularly does not depend on the incident spectrum, but rather depends only on the relative intensity of the different spectral channels as observed by the camera.

However, the specular reflection images  $S_{f,s}$  are not sufficient. For the texture maps *cross-polarized* images  $L_{f,c,s}$  are required for each spectral channel (or, equivalently,  $D_{f,s}$ ). So, using the unpolarized multispectral LEDs of the lighting rig, we capture the unpolarized (“mixed polarization”) image  $M_{f,s}$  for each spectrum  $s$ . An unpolarized lighting image  $M_{l,s}$  for lighting condition  $l$  can be approximated as the sum of cross- and parallel-polarized images:

$$M_{l,s} = L_{l,p,s} + L_{l,c,s} \quad (7)$$

Or equivalently, by substitution:

$$M_{l,s} = D_{l,s} + S_{l,s} \quad (8)$$

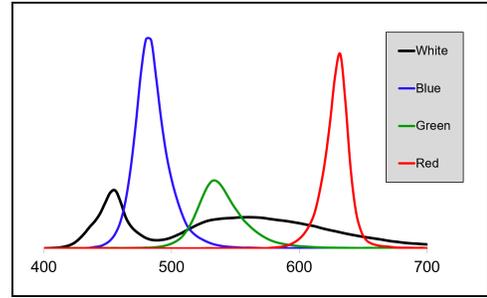
Since we capture images  $M_{f,s}$  and estimate  $S_{f,s}$  for each spectral channel, we can compute  $D_{f,s}$  or equivalently  $L_{f,c,s}$ . The multispectral set of hallucinated images  $D_{f,s}$  provide the diffuse albedo maps required for rendering, after RGB images are formed via color channel mixing. For  $n$  added spectra, we have only added  $n$  *unpolarized* multispectral images to the scan process.

With polarization promotion, we have effectively hallucinated cross- and parallel-polarized images for all spectral channels using only the polarized lighting conditions of one spectral channel and the corresponding unpolarized lighting conditions of the others. Theoretically the polarized spectral channel could be any – polarizing the white LED channel is not a requirement of our approach. However, since the index of refraction has some slight wavelength dependence, comparing specular images under the broad-spectrum white LED with those of the other spectra is advisable to minimize errors caused by the assumption of spectrum-preserving reflections.

## 2 METHOD: MULTISPECTRAL OPTICAL FLOW

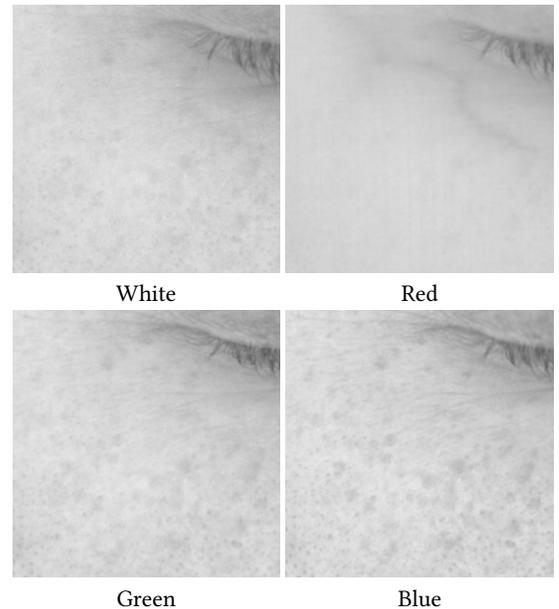
Our optical flow approach is specifically designed to handle incident illumination conditions with similar geometric configurations but different incident spectra.

Formally, our optical flow approach must align  $L_{f,m,s=1..n}$  to  $L_{f,m,0}$ , where  $m$  indicates the mixed polarization condition. Below, we discuss the special case of adding spectral channels comprised of the red, green, and blue LEDs (spectra in Fig. 1).

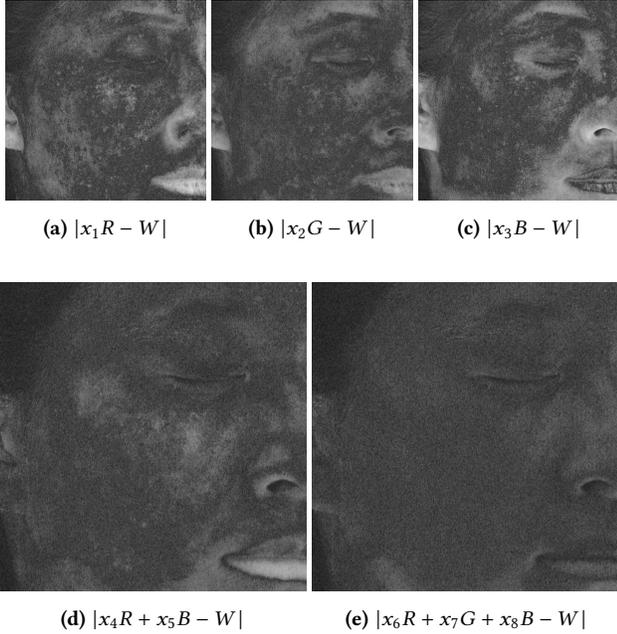


**Figure 1: The spectra of the four LEDs comprising the Light Stage used in this work.**

The appearance of skin illuminated by red, green, and blue light is different owing to the wavelength-dependent effects of sub-surface scattering [Ghosh et al. 2008; Jensen et al. 2001]. When skin is illuminated by a broad-spectrum light source, shallow sub-surface scattered light appears blueish in color, while deeper scattered light appears reddish in color from this wavelength-dependent scattering and light absorption by the skin’s chromophores. For the narrow-band LED illumination, the image under the red LED exhibits less distinct skin texture and a more diffused, soft appearance, in contrast with the image under the blue LED with a great deal of high frequency detail, predominantly from short-wavelength light absorption by epidermal melanin. The image under the green LED is similar to blue, but slightly “softer” (see Fig. 2).



**Figure 2: Inset of facial detail photographed by monochrome camera under different incident illumination spectra, with spectra in Fig. 1. Images have been scaled to the same relative brightness for display.**



**Figure 3: Absolute difference  $e$  in pixel values when approximating the image for white LED, using the red channel  $e_r = 0.00971$ , the green channel  $e_g = 0.00816$ , the blue channel  $e_b = 0.01136$ , the red and blue channel  $e_{rb} = 0.00684$ , and the red, green, and blue channel  $e_{rgb} = 0.00561$  respectively.**

To flow from an image of a subject illuminated by one spectrum to that of a different spectrum, we can naively assume that these images are the same, modulated only by an overall average scale factor  $x_s$  that accounts for the differing fully-spectral modulation of the subject’s average spectral reflectance by the differing incident LED spectra  $s$  and the monochrome camera’s spectral sensitivity. Formally, the assumption is that we can compute  $x_s$  such that  $L_{f,m,s=0} \approx x_s L_{f,m,s}$  for  $s = 1..n$ , approximately satisfying the brightness constancy constraint, so that  $x_s L_{f,m,s}$  may be flowed to  $L_{f,m,s=0}$ . This naive assumption ignores spatially-varying skin spectral reflectance and the effects of sub-surface scattering.

Wilson et al. [2010] defined an iterative optical flow solution to align a pair of complementary images that when added together produced a third target image. The method flowed cross- and parallel-polarized images to mixed polarization images, and flowed spherical gradients and their inverse counterparts towards a full-on even sphere of illumination. We extend the complementary flow of Wilson et al. to the *multispectral* domain, increasing the accuracy of the brightness constancy assumption by combining images across spectral channels. Our key observation is that some linear combination of aligned multispectral images will more closely match the target image,  $L_{f,m,s=0}$ , as compared with each aligned image alone. Fig. 3 demonstrates this effect, where the absolute value of the pixel error is lowest for the linear combination of red, green, and blue images when trying to match the intensity of the white image.

Inspired by the metameric reflectance matching equation of Legendre et al. [2016], in which a target illuminant’s color rendition is matched in a least squares sense by summing the pixel values of color chart squares as illuminated by different LEDs of distinct spectra, we define a least squares procedure to incrementally align spectral channels. We rearrange the pixel values of our monochrome target image  $P_0$  into one column vector  $\vec{P}_0$ . Say we have two column vectors representing pixel values from images  $Q_0$  and  $Q_1$  of different spectral channels to be aligned,  $\vec{Q}_0$  and  $\vec{Q}_1$ . Then we can find values  $x_0, x_1$  that satisfy  $\vec{Q}_0 x_0 + \vec{Q}_1 x_1 = \vec{P}_0$  in a least squares sense. We can then employ complementary flow to align  $Q_0 x_0$  and  $Q_1 x_1$ , which sum to  $P_0$ . In the next iteration, we assume that these two images are correctly aligned, renaming the stacked pixels from the flowed images as  $\vec{N}_0$  and  $\vec{N}_1$ , so we can use them to improve alignment for the next spectral channels. We select new images  $Q_0$  and  $Q_1$  from the set of images not yet aligned, and we compute a least squares solution such that  $\vec{Q}_0 x_0 + \vec{Q}_1 x_1 + \vec{N}_0 x_2 + \vec{N}_1 x_3 = \vec{P}_0$ . Then we can use complementary flow with  $Q_0 x_0$  and  $Q_1 x_1$ , which sum to  $P_0 - N_0 x_2 - N_1 x_3$ , where  $N_0$  and  $N_1$  are already aligned to  $P_0$ . Beyond bootstrapping with an initial target, the algorithm (1) requires no other heuristic. The order of spectral channel alignment is determined automatically, by selecting the next two images that when linearly combined produce the least error as compared with a linear combination of the already aligned images, or in the first iteration,  $P_0$ . In algorithm 1, we define  $N$  as a growing matrix whose columns represent the pixel values of already aligned images that are not  $\vec{P}_0$ . At each step we minimize the term in expression 9:

$$\| [\vec{Q}_0 | \vec{Q}_1 | N] \mathbf{x} - \vec{P}_0 \|^2 \quad (9)$$

The column vectors of Eq. 9 can be linearly independent samples from a color chart lit by each incident spectrum, or pixel values sampled from the actual images after they have been low pass filtered to account for motion since they all have not yet been aligned. Algorithm 1 defines the procedure to apply complementary flow across different illumination spectra. We initialize  $P_0$  as  $L_{f,m,s=0}$  such that the multispectral images align to the spherical gradient images which were captured using the spectrum  $s = 0$ .

### 3 RESULTS AND FIGURES

We show sample images from a monochrome camera facial scan using four spectral channels: red, green, blue and broad-spectrum white (spectra in Fig. 1). Our lighting rig only includes polarizing filters for the white LEDs, so we employ our polarization promotion technique and multispectral optical flow. In Fig. 5, we show input monochrome images and the full-color cross- and parallel-polarized images that can be produced in our pipeline. In Fig. 6 we show a side-by-side comparison of a flash-lit photograph of the subject acquired with a Canon 1DX DSLR camera with a rendering of the subject produced using our monochrome imaging pipeline. For the facial scan, we used 14 monochrome Ximea machine vision cameras, each fitted with a 50mm Fujinon lens and linear polarizer. For the rendering, we used a custom alSurface skin shader and the Arnold global illumination ray-tracer. We tried to match the camera and lighting positions, although in this case the flash-lit photograph of the subject was acquired many days apart from her facial scan. Nonetheless, the subject’s likeness has clearly been captured, and

**Data:**

One target image  $P_0$ , pixels stacked into column  $\vec{P}_0$ ;  
 $n - 1$  unaligned images  $P_{1\dots n}$ , pixels stacked as  $P_{1\dots n}$ ;

**Result:**

$n$  aligned images  $P_0\dots n$ ;

**Procedure:**

Let  $N = []$ ;

Let  $U$  represent the set of unaligned images;

Add  $P_{1\dots n}$  to  $U$ ; Let  $k$  represent the size of  $U$ ;

**while**  $k > 0$  **do****if**  $k > 1$  **then**

Minimize expression 9, solving for  $x$  for all possible combinations of  $\vec{Q}_0$  and  $\vec{Q}_1$  in  $U$ ;

Select  $Q_0$  and  $Q_1$  with minimal error;

Complimentary flow  $Q_0x_0$  and  $Q_1x_1$  to the image formed by unstacking  $\vec{P}_0 - N\mathbf{x}_{2..n}$ ;

Let  $Q_{0,a}$  and  $Q_{1,a}$  represented the aligned images  $Q_0$  and  $Q_1$ ;

Append  $Q_{0,a}$  and  $Q_{1,a}$  as columns to  $N$ ;

Remove images  $Q_0$  and  $Q_1$  from  $U$ ;

**else**

Let  $Q_0 =$  last singleton image in  $U$ ; Let  $\vec{Q}_1 = \mathbf{0}$ ;

Minimize expression 9 solving for  $x$ ;

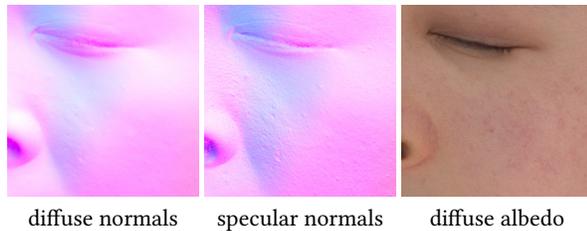
Flow  $Q_0x_0$  to unstacked  $P_0 - N\mathbf{x}_{2..n}$ ;

Remove image  $Q_0$  from  $U$ ;

**end****end**

**Algorithm 1:** Multispectral Optical Flow. Align  $n$  images captured at different time instances and with different spectra.

high resolution facial details are produced along with the color texture map required for rendering. We rendered the subject in Fig. 6 with added skin microgeometry [Graham et al. 2013] to better match the appearance of specular reflections. In Fig. 4, we show a region of the subject's cheek for the single-channel diffuse normal, specular normal, and diffuse albedo texture maps generated from the scan images of Fig. 5. Although our technique produces only a single-channel diffuse normal, the smoother appearance of the diffuse normal map is observed compared with the specular normal map as expected.



diffuse normals    specular normals    diffuse albedo

**Figure 4:** Left: Diffuse normal map for a crop of the cheek region. Center: Specular normal map for the same region. Right: Corresponding diffuse albedo texture map. Each were generated using our monochrome facial scanning pipeline with multispectral polarization promotion.

**ACKNOWLEDGMENTS**

The authors wish to thank Randy Hill, Kathleen Haase, Hao Li, and Christina Trejo for their important support of this work. This project was sponsored by the U.S. Army Research Laboratory (ARL) under contract W911NF-14-D-0005 and in part by a USC Annenberg Ph.D. Fellowship. The content of the information does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

**REFERENCES**

- Abhijeet Ghosh, Graham Fyffe, Borom Tunwattanapong, Jay Busch, Xueming Yu, and Paul Debevec. 2011. Multiview face capture using polarized spherical gradient illumination. *ACM Transactions on Graphics (TOG)* 30, 6 (2011), 129.
- Abhijeet Ghosh, Tim Hawkins, Pieter Peers, Sune Frederiksen, and Paul Debevec. 2008. Practical modeling and acquisition of layered facial reflectance. In *ACM Transactions on Graphics (TOG)*, Vol. 27. ACM, 139.
- Paul Graham, Borom Tunwattanapong, Jay Busch, Xueming Yu, Andrew Jones, Paul Debevec, and Abhijeet Ghosh. 2013. Measurement-Based Synthesis of Facial Microgeometry. In *Computer Graphics Forum*, Vol. 32. Wiley Online Library, 335–344.
- Henrik Wann Jensen, Stephen R Marschner, Marc Levoy, and Pat Hanrahan. 2001. A practical model for subsurface light transport. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*. ACM, 511–518.
- Chloe LeGendre, Xueming Yu, and Paul Debevec. 2016. Efficient Multispectral Reflectance Function Capture for Image-Based Relighting. In *Color and Imaging Conference*, Vol. 2016. Society for Imaging Science and Technology, 47–58.
- Wan-Chun Ma, Tim Hawkins, Pieter Peers, Charles-Felix Chabert, Malte Weiss, and Paul Debevec. 2007. Rapid acquisition of specular and diffuse normal maps from polarized spherical gradient illumination. In *Proceedings of the 18th Eurographics conference on Rendering Techniques*. Eurographics Association, 183–194.
- Cyrus A Wilson, Abhijeet Ghosh, Pieter Peers, Jen-Yuan Chiang, Jay Busch, and Paul Debevec. 2010. Temporal upsampling of performance geometry using photometric alignment. *ACM Transactions on Graphics (TOG)* 29, 2 (2010), 17.



**Figure 5:** From left to right: lighting conditions  $l$ , monochrome parallel-polarized images, monochrome cross-polarized images, monochrome polarization difference images, colorized hallucinated parallel-polarized images, and colorized hallucinated cross-polarized images, for a female subject.



**Figure 6:** Left: Color photograph of a female subject under a flash-lit condition. Center: Rendering of the same female subject from the monochrome scan, photographs in Fig. 5. Note that the scan of the subject and her photograph were completed several days apart, with different cameras. Right: Rendering to show captured geometry without color detail.