Efficient Multispectral Facial Capture With Monochrome Cameras

Chloe LeGendre, Kalle Bladin, Bipin Kishore, Xinglei Ren, Xueming Yu, Paul Debevec
USC Institute for Creative Technologies, Playa Vista, CA, USA

Abstract

*We propose a variant to polarized gradient illumination facial scanning which uses **monochrome** instead of color cameras to achieve more efficient and higher-resolution results. In typical polarized gradient facial scanning, sub-millimeter geometric detail is acquired by photographing the subject in eight or more polarized spherical gradient lighting conditions made with white LEDs, and RGB cameras are used to acquire color texture maps of the subject's appearance. In our approach, we replace the color cameras and white LEDs with monochrome cameras and multispectral, colored LEDs, noting that color images can be formed from successive monochrome images recorded under different illumination colors. While a naive extension of the scanning process to this setup would require multiplying the number of images by number of color channels, we show that the surface detail maps can be estimated directly from monochrome imagery, so that only an additional n photographs are required, where n is the number of added spectral channels. We also introduce a new **multispectral optical flow** approach to align images across spectral channels in the presence of slight subject motion. Lastly, for the case where a capture system's white light sources are polarized and its multispectral colored LEDs are not, we introduce the technique of **multispectral polarization promotion**, where we estimate the cross- and parallel-**polarized** monochrome images for each spectral channel from their corresponding images under a full sphere of even, **unpolarized** illumination. We demonstrate that this technique allows us to efficiently acquire a full color (or even multispectral) facial scan using monochrome cameras, unpolarized multispectral colored LEDs, and polarized white LEDs.*

Introduction

Creating high-quality digital human characters, particularly those based on the likeness of real people, is a long-standing goal in computer graphics, with applications in films, video games, simulations, and virtual reality. Computational imaging and illumination systems have been developed to faithfully capture a subject's facial shape and appearance to produce highly photo-realistic renderings of the subject's digital double. Such systems can be broadly grouped into two categories: (1) those using even, diffuse illumination and multiple camera viewpoints to estimate facial geometry using multiview stereo approaches, and (2) those using multiple camera viewpoints in combination with specialized illumination patterns, generating facial geometry of a comparatively higher resolution using variants of shape from shading or photometric stereo for surface normal estimation [43].

For the methods of the second category, such as that of Ma et al. [32] and subsequent approaches [17, 40], polarization filters are placed in front of the lights and the cameras, and polarization difference imaging separates the sub-surface and specular reflections from images for each gradient illumination condition, exploiting the polarization-preserving aspect of specular reflections from the skin. Ma et al. [32] demonstrated that separating sub-surface and specular reflections improved 3D reconstruction, since surface normals estimated from the sharp, unblurred specular reflections more accurately captured the pore-level details of facial geometry. Additionally, the cross-polarized images provided a better representation of the subject's colored diffuse albedo, recording light that has traveled through the skin, having undergone multiple sub-surface scattering events and therefore depolarization before reflecting back towards the camera.

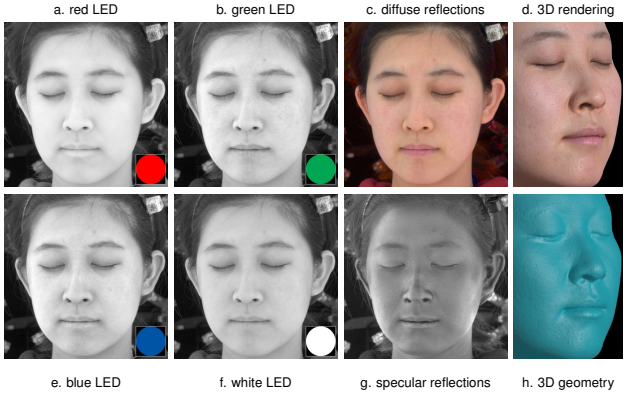


Figure 1: **a,b,e,f:** Monochrome photographs of a subject lit by red, green, blue, and white LEDs; **c:** Colorized diffuse reflection image produced by mixing a,b,e, and f; **g:** Monochrome polarization difference image showing specular reflections; **d** Full-color rendering of the subject; **h:** Geometry rendering of the subject with no diffuse albedo. Renderings are produced using only monochrome images.

Since a color diffuse albedo texture map is required for rendering a photo-realistic digital double, most facial scanning systems of both categories [4, 5, 32, 17, 40] use tristimulus RGB cameras and broad-spectrum white light sources for acquisition. One recent exception is the work of Fyffe et al. [14], where monochrome machine vision cameras were used in combination with unpolarized LED-generated blue light. This shorter wavelength light revealed more texture cues in the skin for multiview stereo and optical flow computations. However, this single-shot approach only generated facial geometry and did not capture reflectance information or produce the colored texture maps needed to produce photo-realistic renderings.

Photography with monochrome cameras offers a few obvious theoretical advantages. First, for a color camera, the light-absorbing color filters placed in front of the imaging sensor may absorb more than two thirds of the total incident light per pixel, requiring a light-boosting compromise somewhere else in the system across the exposure triad (increasing sensor gain, exposure time, or aperture size, none of which are desirable for high-resolution facial scanning), or the use of even brighter lights, which could be uncomfortable for live subjects. Additionally, most modern color cameras include RGB filters in the color filter array (CFA) arrangement of a typical Bayer pattern, reducing the effective image resolution for each color channel. Full-color images are produced using various up-sampling methods (referred to as debayering or demosaicing algorithms). In contrast, monochrome cameras allow for imaging with comparatively less incident light at a higher true image resolution, while only generating a single-channel intensity image. In this work, our goal is to gain the aforementioned benefits of monochrome camera image acquisition in a high-resolution facial scanning system, without sacrificing the color information that is required for rendering.

While color images are most often produced using tristimulus RGB cameras with spectral sensitivity functions similar to the human visual system, they can also be produced using active illumination – by sequentially illuminating a subject with at least three differently colored light sources, capturing images using a *monochrome* sensor with a broad spectral response. This technique trades the spatial multiplexing of typical RGB cameras for temporal multiplexing, increasing the image resolution for each color channel at the expense of requiring more photographs. The resulting color images will be similar if the power spectra of the narrow band light sources (modulated by the monochrome camera's spectral response curve) are similar to those of the broad-spectrum white light source modulated by a color camera's spectral response curves. LED light sources have been used for this technique, though their emission spectra are often narrower than typical camera spectral sensitivity functions.

While the lighting systems used for high-resolution facial scanning [32, 17, 40] were comprised of broad-spectrum white LEDs, other omnidirectional lighting rigs have been built using red, green, and blue LEDs [8, 22]. These systems were not developed for facial scanning, but systems like these could enable monochrome camera color imaging via time-multiplexed RGB illumination. Furthermore, several omnidirectional *multispectral* light stages have been built [1, 21, 28, 30, 9], incorporating RGB LEDs along with other LEDs of distinct spectra, often with six or more spectral channels per light source. Such systems would enable multispectral image acquisition, rather than just RGB.

The computational lighting-based facial scanning techniques already employ time-multiplexed illumination, requiring at least eight images [32] of a subject under different lighting conditions for a complete scan. At first glance, one might assume that using a monochrome camera and colored LEDs instead of a color camera and white LEDs would require at least 3 times the originally required number of images, assuming three-channel output. Or, in the multispectral case, one might expect to need at least n times as many images, where n is the number of desired spectral channels of the output. Such a large number of images would be impractical to acquire for live subjects, where scan time should be minimized to reduce the negative effects of slight subject motion across frames.

However, in this work, we evaluate for each stage of the high-resolution facial scanning process, whether or not full color imaging is necessary. From this analysis, we show how to efficiently capture a full color (or multispectral) facial scan using monochrome cameras and colored LEDs, requiring only n additional images, where n is the desired number of added spectral channels used for producing the diffuse albedo texture map. We also introduce a new *multispectral optical flow* approach that corrects for subject motion and chromatic aberrations, based on the method of *complementary flow* [40].

Additionally, for the case where a system's broad-spectrum white LEDs are already polarized for facial scanning while its multispectral colored LEDs are not, we introduce an approach that we call *multispectral polarization promotion*, which allows us to hallucinate cross- and parallel-polarized images of a subject for each spectral channel even when all but one are unpolarized. We compute the per-pixel amount of light reflected specularly relative to the quantity of incident light for a given lighting condition using the polarized white LEDs and polarization difference imaging. Since skin is a dielectric, the proportion of incident light reflected specularly should mostly not depend on the spectrum of the incident illumination. Therefore, the specular reflection image can be approximated as consistent across the other spectral channels, up to a scale factor accounting for the monochrome camera's potentially different sensitivities to each of the spectral channels or differing LED intensities. With this information, for each spectral channel we can hallucinate its cross-polarized image of sub-surface scattered light by subtracting the hallucinated per-pixel specular reflection image from a captured photograph of the subject under an unpolarized lighting condition for that incident spectrum.

In summary, our contributions are:

- We demonstrate that full-color or multispectral high-resolution facial scans using state-of-the-art techniques [32, 17, 40] can be efficiently acquired using monochrome cameras and colored LEDs instead of color cameras and broad-spectrum white LEDs, thereby increasing image resolution and reducing light requirements.
- We show that such a scan can be acquired by adding only n photographs to the normal scan sequence, where n is the number of added spectral channels.
- We introduce a new multispectral complementary flow approach to align images captured under illuminants of different spectra.
- We show that polarized light sources are only required for one spectral channel of the illumination system, and for the remaining spectral channels we can promote unpolarized images to cross- and parallel-polarized as required.

Related Work

3D Facial Capture

There are several related goals in computer graphics that require high-resolution facial scanning techniques. For some applications, it may be desirable to capture a live actor's performance and then have the ability to replay the performance from a new camera or viewpoint, or under new illumination. Active illumi-

nation techniques have been proposed to synthesize the required new views and the subject's appearance under novel lighting using image-based relighting approaches [7, 38], although these are data-intensive captures requiring on the order of hundreds of photographs per scan.

Another goal is to enable computer graphics artists to produce an entirely novel performance for a given actor by animating the actor's digital double. For photo-realistic visual results, high-resolution skin geometry and reflectance information are required, and so static high-resolution scans of the subject in a variety of extreme facial expressions are captured. After a topological registration step, animators can mix or cross-dissolve between these facial expressions to create a new animation [31]. The static facial scans are often generated using computational illumination techniques [32, 17], which we extend for the monochrome camera case in this work.

Ma et al. [32] introduced one such system using a series of eight lighting patterns designed to produce surface normal estimates. These patterns were produced using a Light Stage, a spherical lighting rig comprised of broad-spectrum, white LEDs. Noting the symmetry of bidirectional reflectance distribution functions (BRDFs) around the surface normal or reflected light directions for diffuse and specular BRDFs respectively, they derived per-pixel estimates for these directions by photographing the subject lit by a sequence of polarized spherical, linear gradient lighting conditions. Intuitively, these lighting conditions produced estimates of the centroids of the BRDFs, yielding *diffuse and specular normals*.

Since the high-resolution capture methods [32, 17] are limited to static facial poses, significant research has endeavored to capture dynamic human facial performance. Such captures allow for an actor's performance to be transferred to a new digital character or allow the actor's performance to be re-rendered in some other novel way. Multi-camera video recording setups enable dynamic multiview stereo based 3D reconstructions [11, 4, 5, 14]. Since these techniques only require an image of the subject in a single, even lighting condition per viewpoint, they can be applied to video sequences of an actor's performance. However, such methods do not capture skin reflectance information, and they estimate high resolution geometric details using the technique referred to as "dark is deep," where dark pixels are assumed to represent surface concavities. This assumption breaks down for the commonly-occurring cases of skin pigmentation or blemishes, where dark pixels are the result of light absorption rather than shadowing. As such, multiview computational illumination approaches [17, 32], though requiring multiple input photographs of the subject under different time-multiplexed lighting conditions, still provide superior geometric accuracy. Dynamic scanning methods can be augmented or initialized by one or more high resolution static scans [3, 2, 13, 25, 16], providing an additional application for our monochrome imaging approach.

To capture the details of a high-resolution scan for dynamic facial performances, Wilson et al. [40] extended the work of Ma et al. [32] by introducing a novel optical flow based approach. They added gradient illumination conditions that were complementary to those of Ma et al. [32], such that pairs of images of a subject under different lighting conditions could be added together to produce an image of the subject as lit by a full sphere of even illumination, enabling temporal image alignment by satisfying the brightness constancy constraint of optical flow. This flow-based image alignment also improved the quality of static scans, offering superior image alignment and better signal-to-noise ratios as an over-complete image basis was acquired. The monochrome facial scanning approach that we describe is applicable to this technique as well, specifically as we define a custom optical flow step that aligns images captured under different illumination spectra.

Polarization-based Analysis of Reflected Light

When light reflects from a dielectric surface such as human skin, the specular component can be polarized while the diffuse reflection is largely unpolarized. This difference has been leveraged in the classic computer vision literature [42, 41, 34] to aid in the separation of diffuse and specular reflections. Debevec et al. [7] actively polarized a light source and employed polarization difference imaging to separately model diffuse and specular reflectance lobes of a human face. The previously described high-resolution facial scanning approaches [32, 17, 40] polarize spherical lighting conditions to estimate per-pixel surface normals from each reflectance component. Color space analysis may also be used for diffuse-specular separation [34, 33], which Fyffe et al. [12] extended for facial scanning without polarization filters. In another filter-free approach, Tunwattanapong et al. [37] used continuous lighting patterns of higher order spherical harmonics to separate diffuse and specular reflections as well as to estimate specular roughness and anisotropy. In this work, we use polarization difference imaging with monochrome cameras for diffuse-specular separation, but other filter-free approaches would be of interest for future work.

Multispectral Illumination and Reflectance Measurement

Multispectral LEDs have been used in recent years for time-multiplexed lighting, due to the advancement of solid state lighting technology. One application has been for spectral reflectance estimation of materials, which can be performed using a small LED array [27, 35, 19, 36] or a multi-source rig capable of producing light from many directions [1, 21, 28]. Multispectral LEDs have also been used for lighting reproduction, where the goal is to reproduce a particular illuminant's power spectrum or match its color rendition properties in a studio environment. This technique has been practiced again for a single LED array [39, 10] or for an omnidirectional lighting rig [30, 9]. Such a rig has also been used for multispectral image-based relighting [29], extending to the multispectral domain the work of Debevec et al. [7], in which a subject's appearance in any lighting environment can be produced by summing together its appearance as illuminated by a set of basis lighting conditions. Our objective is different from each of these previous works, as we seek to generate a high-resolution 3D facial model with multispectral textures. However, our work uses the image-based metameric reflectance matching technique of LeGendre et al. [29, 30] to produce color images from a set of multispectral basis images.

Multispectral light sources have been used directly for 3D reconstruction as well, extending photometric stereo [43] to a single-shot and therefore video-rate approach by trading temporal-multiplexing for spectral multiplexing [23]. Fyffe et al. [15] extended this approach to the dynamic 3D reconstruction of faces, although they used cross-polarized images and assumed

Lambertian skin reflectance, limiting the resolution of the recovered 3D geometry.

Analysis and Equations

In this section, we analyze which parts of a high-resolution facial scanning pipeline can be achieved with monochrome cameras alone, and which parts require the addition of multispectral LEDs to obtain color information. In our analysis, we consider a monochrome camera extension to the scanning approach of Ghosh et al. [17], which incorporates the gradient illumination patterns of Ma et al. [32] and the optical flow techniques of Wilson et al. [40], but with a spherical polarization scheme that enables diffuse-specular reflection separation across multiple camera viewpoints. We also justify our technique of multispectral polarization promotion and introduce a method for multispectral optical flow.

Monochrome Multiview Stereo

Initially, a low-resolution 3D reconstruction of the face is generated using passive multiview stereo. Images of the subject are captured under a single even, diffuse lighting condition from a variety of viewpoints. Multiview stereo approaches do not require RGB images; they even operate more efficiently when using only intensity information (RGB data converted to gray-scale), extracting cross-view correspondences on one third of the input data. As noted by Fyffe et al. [14], skin texture cues are highly visible under shorter wavelength blue light, which is absorbed by spatially-varying skin pigmentation (see Fig. 3). Thus, color images are trivially not required for multiview stereo, and blue illumination with monochrome cameras will improve stereo matching.

Monochrome Specular Normals

Ma et al. [32] describe spherical gradient illumination conditions used to infer surface normals, for surfaces that primarily reflect light diffusely (Lambertian) and specularly. They introduced "specular normal maps," showing that polarization difference imaging combined with gradient illumination conditions could yield geometry with resolution comparable to that achieved with laser scanning. Cross-polarized images were subtracted from parallel-polarized images, producing specular reflections image for each gradient lighting condition, from which the specular normals were derived.

Since skin is a dielectric, specularly-reflected light is mostly of the same spectrum as the incident illumination. Fresnel's equations describe the amount of light incident to an interface that is reflected versus refracted, depending on the indices of refraction of the interface materials. For typical dielectric materials, the index of refraction has some dependency on wavelength, but this dependency is very slight across the visible spectrum, and computer graphics practitioners commonly assume that reflectivity is not a function of wavelength. This is a reasonable assumption for dielectric materials. Accordingly, the specular reflection image of a face produced via polarization difference imaging is largely "colorless" (see Fig. 2). Therefore, we do not need color cameras to compute "specular normals." This makes sense intuitively, since specular normals encode geometry rather than color. Therefore, to compute specular normals, we could theoretically generate polarized gradient illumination conditions using any illuminant to which a monochrome camera has some sensitivity.

So far, both 3D reconstruction steps of a high-resolution facial scan, the multiview stereo coarse reconstruction followed by the specular reflection based surface normal computation, do not require color images. If we only wanted to measure high-resolution facial geometry without computing the subject's diffuse albedo texture map, then a monochrome camera scan would require exactly the same number time-multiplexed lighting conditions as a color camera scan.

Figure 2: **Left:** Cross-polarized color photograph of a subject under an even sphere of white light (sRGB). **Right:** Specular reflection image computed via full color polarization difference imaging. The image has been white-balanced to the color of the white light source. Specular reflections are largely colorless.

Diffuse Reflectance

However, to render a subject's digital double in color, artists require a colored texture map of the subject's diffuse reflectance, which is approximated using a view-dependent synthesis of color images of the subject illuminated by a full sphere of cross-polarized white light. If our light stage includes *polarized colored LEDs* in the same polarization arrangement as the white LEDs of Ghosh et al. [17], then we could capture a cross-polarized image of the subject under a full even sphere of illumination for each of the available spectral channels, producing the images required to generate a multispectral diffuse texture map of the subject. This multispectral diffuse texture map could then be used to generate an RGB texture map of the subject's appearance under a particular illuminant (see subsection: Color Channel Mixing).

Note that up to this point, despite capturing the spherical gradient images with a monochrome camera, the only additional images that must be captured are those of a subject under a full sphere of cross-polarized illumination for each added spectral channel. For n added spectra (excluding the white LED, for which we already have the full sphere cross-polarized condition), we have only added n images to the scan process. Importantly, we do not need to capture the full polarized gradient illumination sequence for each spectral channel (or even with just RGB LEDs) to obtain a high-quality scan complete with the subject's color or multispectral diffuse albedo.

Multispectral Polarization Promotion

However, polarizing all the colored LEDs of a lighting rig not only adds complexity, but it also absorbs over half of the light emitted by the LEDs. Since colored LEDs are often used for applications like live-action compositing with lighting reproduction [8, 22, 30] where video-rate recording demands short ex-

posure times, halving the light output of the colored LEDs is undesirable. As an alternative, 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. [17].

For clarity, we extend the variable naming conventions of Ma et al. [32]. 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 [32] 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 $L_{l,c,s} = \frac{1}{2}D_{l,s}$, 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 $L_{l,p,s} = \frac{1}{2}D_{l,s} + S_{l,s}$. Therefore, for each gradient lighting condition l and spectrum s, the specular reflection image $S_{l,s}$ is produced via polarization differencing:

$$S_{l,s} = L_{l,p,s} - L_{l,c,s} \tag{1}$$

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 $C(\lambda)$:

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

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 (ρ_{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:

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

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:

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

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

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

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:

$$\frac{S_{l,0}}{W_0} = \frac{S_{l,s}}{W_s} \tag{6}$$

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:

$$M_{l,s} = L_{l,p,s} + L_{l,c,s} \tag{7}$$

Or equivalently, by substitution:

$$M_{l,s} = D_{l,s} + S_{l,s} \tag{8}$$

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. Again, 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.

Monochrome Diffuse Normals

Ma et al. [32] introduced "diffuse normals," which could be used in a "hybrid normal" shader to simulate the effects of sub-surface scattering in a real-time rendering application. In our multispectral polarization promotion approach, we only photograph diffuse gradient lighting conditions for a single monochrome image channel for the white LED, so we generate single-channel diffuse normals rather than typical RGB diffuse normals. In our results section, we visualize the colorized spherical gradient illumination images by considering that the ratio of a cross-polarized gradient condition to a full sphere cross-polarized condition should be approximately the same across spectral channels. After photographing $L_{l,c,0}$, $L_{f,c,0}$, and hallucinating $L_{f,c,s}$ for each spectral channel, we can approximately compute $L_{l,c,s}$ for lighting conditions l of gradients x, y, z:

$$\frac{L_{l,c,s}}{L_{f,c,s}} \approx \frac{L_{l,c,0}}{L_{f,c,0}} \tag{9}$$

This is an approximation that does not consider the wavelength-dependent optical properties of the skin demonstrated through diffuse normals computation.

Color Channel Mixing

Once the diffuse or sub-surface scattered reflection images for each spectral channel have been photographed or computed via polarization promotion, we generate an RGB image of the diffuse reflections using image-based multispectral metameric reflectance matching as in LeGendre et al. [30], extended to the monochrome imaging case [29]. The weights of the different spectral channels, α_s must be computed separately for each desired output color channel, yielding $3s$ degrees of freedom. The pixel values of a color chart for a given target illuminant are represented by pixel values P_{jc} where j is the index of the given color chart patch and c is the target camera's c'th color channel. N_{js} is the average pixel value of color chart square j under LED spectrum s, and $\mathbf{N}$ is the $j \times s$ matrix whose columns correspond to the LED spectrum s and whose rows correspond to the color chart square j. Eq. 10 should be minimized for each color channel of the target color chart, producing in our case three α vectors, which are the lighting primaries. This locks in the RGB diffuse albedo to the lighting condition illuminating the target chart and specific camera. Optionally, a color matrix may also be applied.

$$\sum_{j=1}^{m} (P_j - \sum_{s=1}^{n} \alpha_s N_{js})^2 = ||\mathbf{P} - \mathbf{N}\alpha||^2 \tag{10}$$

Optical Flow

As in other high-resolution facial scanning approaches [40, 17, 32], temporal alignment between photographs is a prerequisite for all computations requiring more than one pattern, i.e. estimation of diffuse and specular normals, polarization promotion, and color channel mixing. For the monochrome imaging approach, we need to flow across spectral channels, not only to account for potential movement between frames but also to correct for chromatic aberrations. Formally, our optical flow approach must additionally 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. 5).

The appearance of skin illuminated by red, green, and blue light is different owing to the wavelength-dependent effects of sub-surface scattering [18, 26]. 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. 3).

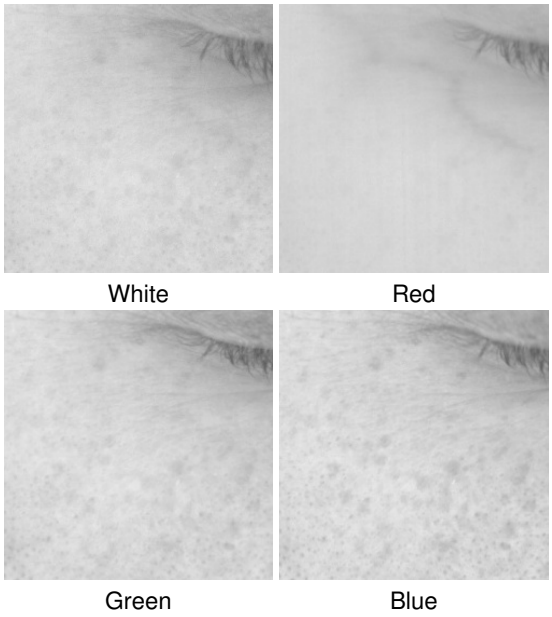
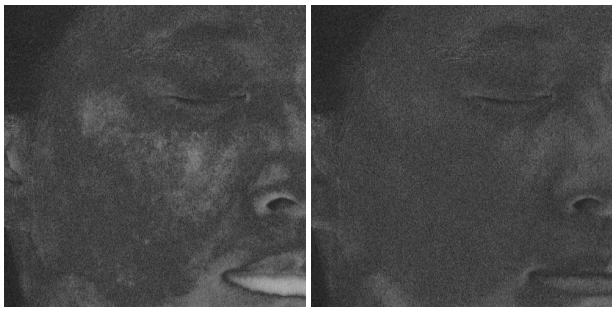


Figure 3: Inset of facial detail photographed by monochrome camera under different incident illumination spectra, with spectra in Fig. 5. Images have been scaled to the same relative brightness for display.

To flow from an image of a subject illuminated by one spectrum to that of a different spectrum, we can naïvely 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

(a) $|x_1R-W|$ (b) $|x_2G-W|$ (c) $|x_3B-W|$



(d) $|x_4R+x_5B-W|$ (e) $|x_6R+x_7G+x_8B-W|$

Figure 4: 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.

such that $L_{f,m,0} \approx x_s L_{f,m,s}$ for $s = 1..n$, approximately satisfying the brightness constancy constraint, so that $x_s L_{f,m,s}$ may be flowed to $L_{f,m,0}$. This naive assumption ignores spatially-varying skin spectral reflectance and the effects of sub-surface scattering.

Wilson et al. [40] 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, $L_{f,m,0}$, as compared with each aligned image alone. Fig. 4 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 expression (Eq. 10) and complementary optical flow [40], we define a least squares procedure to incrementally align images of the same subject captured as illuminated by different spectral channels. For a set of two or more unaligned images, we compute the amounts x_s of each image $L_{f,m,s}$ for $s = 1..n$, that, when all added together, best produce the target image $L_{f,m,0}$, as in Eqn. 11.

$$\underset{\mathbf{x}\geq 0}{\arg\min}(||L_{f,m,0} - \sum_{s=1}^{n} x_s L_{f,m,s}||^2) \tag{11}$$

We scale the unaligned images by these amounts x_s. Then, as in Wilson et al. [40], we initialize the flow fields for each scaled image to 0, indicating no motion, and iteratively update flow estimates for each spectral channel. During the first step of one iteration, the flow field for an unaligned image is estimated by assuming that the other unaligned images' flow fields are constant. In the the next step of the iteration, we estimate the flow field for a different unaligned image, while the flow fields for other images including the scaled aligned image of the previous step are assumed constant.

To find x_s values, we could use linearly independent samples from a color chart lit by each incident spectrum, or pixel values sampled from the actual images after low pass filtering to account for motion, as they are initially unaligned. In practice, we sample pixel values from filtered images and use a non-negative least squares solver. However, the solver may suggest $x_s = 0$ for some spectrum s, which means that the corresponding image will never be aligned during the complementary flow. A weight of $x_s = 0$ in the solve means that adding this spectrum does not further help to minimize color error, which implies that its spectral contribution is either redundant or not useful due to a lack of spectral overlap with the target illuminant. To handle redundancy, for each source image with $x_s = 0$ in the initial solve, we can subsequently find the set of already aligned images that, when combined, best match it in a least squares sense. Then, we can compute simple optical flow to align the source image to the already aligned linear combination target. We note that this is different from the iterative complementary step, as the color mixing solve endeavors to best match an unaligned *source* image $L_{f,m,s}$ rather than the *target* image $L_{f,m,0}$, and furthermore this step does not use the complementation constraint. This final step ensures that every image is ultimately aligned to the target.

To further improve the robustness of our multispectral optical flow, we note that negative pixels may be produced when linearly combining images across spectral channels, even when using a non-negative least squares solve. As an example, during one iteration, say we solve for the amount of a red image R and a green image G that best produce the target white image W. Then, the flow step would alternate between aligning the source image R to the target image $W-G$ and the source image G to the target image $W-R$, possibly producing some negative pixel values. Rather than clamping the images to zero, we find the minimum and maximum pixel values of both the source and target images, and remap both to the same pixel value range. This changes the relative intensity of both images, so we further normalize each remapped image by dividing each by a highly smoothed, Gaussian-blurred version of itself.

Results and Discussion

Monochrome Camera Facial Scan

First, we show sample images from a monochrome camera facial scan using four spectral channels: red, green, blue and broad-spectrum white (spectra in Fig. 5). The lighting rig used for this facial scan only has polarizing filters for the white LEDs, so we employ our polarization promotion technique and multispectral optical flow. In Fig. 7, we show input monochrome images and the full-color cross- and parallel-polarized images that can be produced in our pipeline. In Fig. 8 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 pro-

duced using our monochrome imaging pipeline. For the facial scan, we used 14 monochrome Ximea xiQ MQ042MG-CM 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 high resolution facial details are produced along with the color texture map required for rendering. For the renderings in Fig. 8, in keeping with the state-of-the-art, we added image-based skin microgeometry [20] to improve the appearance of specular reflections. In Fig. 6, 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. 7, without added microgeometry. 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.

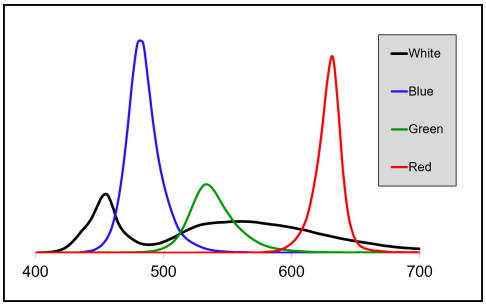


Figure 5: The spectra of the four LEDs comprising the light stage used in this work.

Figure 6: **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.

Comparison with Color Imaging

To demonstrate the improvement in the resolution of geometric "specular normals" when using a monochrome camera as compared with a color camera, we photographed a subject using two cameras with the same sensor, but one color and one monochrome. The two machine vision cameras, a Ximea xiC MC124CG-SY (color) and a Ximea xiC MC124MG-SY (monochrome) both use the Sony IMX253 sensor. They were placed immediately adjacent to one another, though not on the exact same optical axis. With both cameras, we photographed a subject illuminated by polarized spherical gradient illumination conditions in the lighting rig, aligned the images with optical flow, and then computed the "specular normals" of Ma et al. [32]. We show the results in Fig. 9. In the top row, we compare the specular normals obtained using a monochrome camera with those obtained using a color camera, using the adaptive homogeneity-directed demosaicing algorithm [24] for obtaining color images. In the bottom row, we compare the specular normals obtained using a monochrome camera with those obtained using a color camera and a simple linear interpolation demosaicing algorithm. In both cases, the specular normals computed with the monochrome camera images are sharper and show a greater level of detail.

Multispectral Polarization Promotion

Next, to validate our new technique for hallucinating cross- and parallel-polarized multispectral images when only one spectral channel is polarized, we photographed a subject in a *different* lighting rig where all four spectral channels (red, green, blue, and broad-spectrum white) are polarized. We were therefore able to generate "ground truth" polarization difference images for each spectral channel. For this experiment, we used a *color* Ximea xiC MC124CG-SY camera, fitted with a linear polarizer and 50 mm Schneider lens (though only a monochrome camera is required). We photographed the subject under each spectral channel for the cross and parallel polarization states, producing the color polarization difference images in the top row of Fig. 10. Next, we simulated a monochrome camera response for the polarization difference images, converting the color images to grayscale using $Y = 0.2126R + 0.7152G + 0.0722B$. We calibrated the relative LED intensities across spectral channels as observed by the camera, photographing a color chart and measuring the pixel values of the white square, and we then scaled the monochrome polarization differences for each spectral channel according to this calibration. These scaled images are shown in the middle row of Fig. 10. They all appear to be visually the same, validating the assumption that the amount of light reflected specularly from the skin does not depend on the incident illumination spectrum. Indeed, for incident illumination of pixel intensity 1.0, the average absolute difference between the polarization difference images for the white LED as compared with the red, green, and blue LEDs are $0.0160, 0.0016$, and 0.0012 respectively, sampled along a large section of the subject's cheek region. These absolute difference images are shown in the bottom row of Fig. 10.

Multispectral Optical Flow

Next, to evaluate our optical flow technique, we aligned the red, green, and blue LED images of our scan subject, in Fig. 1(a), (b), and (e), to the white LED, Fig. 1(f). For these images, we found that even naïvely flowing a weighted version of each spectral channel directly to the white image worked well in practice and visually aligned the images. We expect that naïve optical flow was successful for a number of reasons. First, the reflectance spectra of skin are relatively smooth, and the spectrum of our target white LED covered a large portion of the visible wavelength range and overlapped with each of the other LED spectra (see Fig. 5). Additionally, the blue and green images revealed high frequency details required for optical flow approaches, allowing them to be easily aligned to white after global intensity matching followed by local intensity normalization. The red image lacked high frequency detail, showing a large texture-less region for the

Figure 7: 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.

skin, a challenging input for most optical flow techniques. However, if a pixel is misaligned in a texture-less region, the result can still be visually acceptable, since it will take on a nearby pixel's nearly identical intensity. Finally, each image was captured with a 12 millisecond exposure time, and therefore a complete facial scan took less than one quarter of a second. With a short capture time and the subject trying not to move, motion across frames was very slight, on the order of less than 3 pixels.

Though we found our complementary multispectral optical flow technique was not required for the input images in Fig. 1, we captured an additional scan subject to demonstrate the utility of our approach for the challenging case of photographing objects with diverse reflectance spectra. We painted a Styrofoam mannequin head with different brightly colored paints (see Fig. 11a) and photographed it under slight rigid motion inside the lighting rig with a Ximea monochrome xiQ MQ042MG-CM camera, using the same four spectral illumination conditions (red, green, blue and white LEDs), while adding amber and cyan LEDs (spectra in Ref. [30]). In Fig. 12, we compare results for four different multispectral optical flow approaches, where all four use the same core optical flow algorithm, parameters, intensity normalization techniques, and non-negative least squares solver. Fig. 12 visualizes the magnitude and direction of the computed flow fields as saturation and hue respectively. As the mannequin only moves rigidly, the pixel values of these flow visualizations should be smoothly varying within a spectral channel, indicating a similar direction of motion for all image pixels. To demonstrate the appearance of a smoothly varying flow field, we trivially aligned a white LED image to a different white LED image (Fig. 11c), and visualized the flow result in Fig. 11b. The four multispectral optical flow approaches with results in Fig. 12 are the following:

- **Naïve Flow**: Separately flow a scaled version of each spectral channel to white.
- **Single-Image Incremental Flow**: First, find the image that, when scaled, best matches white in a least squares sense, and flow this scaled image to white. Then flow each subsequent image individually to a linear combination of all previously aligned images, including white. (This uses metameric reflectance matching but *not* complementary flow.)
- **Two-Image Complementary Flow**: First, find the *pair* of images that, when combined, best match white in a least squares sense, and apply two-image complementary flow

Figure 8: **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. 7. 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. For both renderings, we applied image-based skin microgeometry [20] to the displacement maps.

to iteratively align both scaled images to white. Then flow each subsequent image individually to a linear combination of all previously aligned images, including white. (This uses metameric reflectance matching and *two-image* complementary flow.)

- **Multi-Image Complementary Flow**: Find the linear combination of *all* of images that, when combined, best match white in a least squares sense, and apply multi-image complementary flow to iteratively align *all* images to white. If the initial solve does not use all spectral channels, flow any unused images individually to a linear combination of all previously aligned images, including white. (This uses metameric reflectance matching and *multi-image* complementary flow.)

Visually, our results demonstrate that the two-image complementary approach yields more smoothly-varying flow fields for the rigidly moving mannequin than either the naïve or single-image incremental approach, indicating that our complementary multispectral optical flow approach can improve alignment for challenging subjects with diverse reflectance spectra and considerable movement. Though our results are not perfect, as some non-smoothly varying color changes are observable, there is clear improvement over the naïve approach. While the residual color error when approximating white is smaller for the multi-image complementary approach, which might help to better meet the brightness constancy constraint of optical flow, this comes at the cost of introducing a larger number of "guessed" initial flow fields in the complementary flow framework. As a result, qualitatively, we observe the best results for the two-image complementary approach, where fewer initial flow field guesses are required.

For all experiments, we used the OpenCV 3.3 [6] implementation of TV-L1 optical flow [44], with the following parameters: $\tau = 0.25, \lambda = 0.15, \theta = 0.3$, nscales=5, warps=5, $\varepsilon = 0.01$, innerIterations=30, outerIterations=10, scaleStep=0.5, $\gamma = 0$, and no median filtering. For images of size 2048 x 2048, we used a Gaussian blurring kernel of width $= 21$ for both the metameric reflectance solve and the intensity normalization.

Color Mixing for Diffuse Albedo

Another benefit of our multispectral scanning approach is that we can approximate the color rendition properties of different illuminants as observed by a particular camera when generating cross-polarized images and diffuse albedo texture maps (see bottom row of Fig. 13), using Eqn. 10. It is well known that simply applying a 3×3 color channel mixing matrix to an RGB image cannot correct for spectral differences between illuminants. We therefore evaluate our technique's ability to match the color rendition properties of three different illuminants (daylight, fluorescent, and tungsten) as observed by a Canon 1DX camera. With the same camera, we photographed a color chart lit by white LEDs. We then computed the best 3×3 color matrices to try to "correct" this image to match the color rendition properties of different real-world illuminant images. We repeated this test for a different Ximea color camera as well. With a monochrome Ximea camera, we then captured the appearance of the color chart lit by six LEDs of distinct spectra, WRGB plus cyan (C) and amber (A), generating the $\mathbf{N}$ matrix of Eq. 10. We solved for the multispectral image primaries for each illuminant when using WRGBCA, WRGB, and RGB LEDs only. We compare the color rendition of each approach in Fig. 13. The background squares represent the target colors, while the foreground dots represent the colors

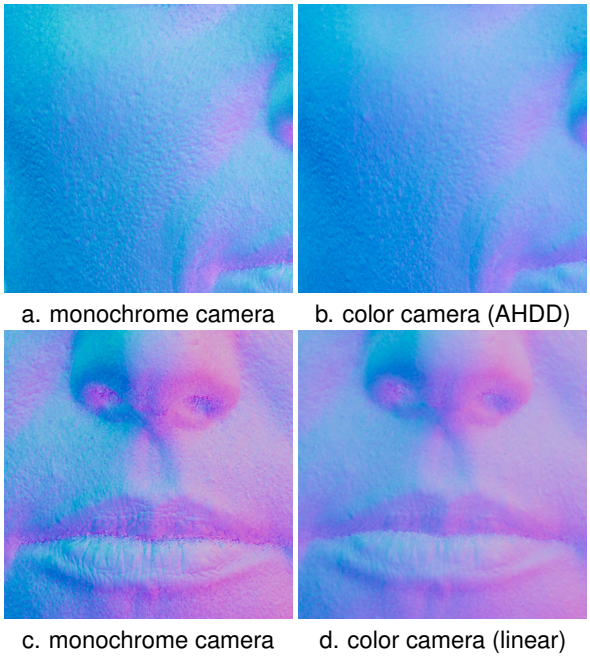


Figure 9: **a:** Specular normals of a subject's cheek, computed using polarized gradient illumination with a monochrome camera, and **b:** with a color camera, using the adaptive homogeneity-directed demosaicing algorithm [24] to obtain color images. **c:** Specular normals of a subject's lip region, again computed using polarized gradient illumination with a monochrome camera, and **d:** with a color camera, using simple linear Bayer pattern interpolation to obtain color images. Even when using a color camera in combination with a sophisticated demosaicing algorithm, photography with monochrome cameras yields higher resolution geometry, as observed via the sharpness of skin details in **a** and **c**.

achieved by applying the color matrices (top two rows) or by mixing spectral channels (bottom chart row). When the dots "disappear," it means we have achieved good color rendition. The multispectral approach with six spectral channels visually and quantitatively out-performs the color-matrix approach for daylight and incandescent light, especially for the important skin-colored square of the color chart. Color rendition can be improved with multispectral monochrome imaging with six spectral channels, which suggests that the skin's wavelength-dependent sub-surface scattering effects maybe better estimated using our approach, if an artist knows ahead of time the color rendition properties of a virtual scene's dominant illuminant. We show that generating texture maps with WRGB alone with monochrome cameras does *not* out-perform the color matrix approach, however it performs significantly better than RGB only.

Future Work

In this work, we have used polarization difference imaging for diffuse-specular separation. It would be of interest to evaluate other methods for this step as well, particularly an approach employing higher order spherical harmonic lighting conditions [37] instead of polarizers. Ideally, polarizing filters could be avoided for all spectral channels to maximize light output and further reduce exposure times. Furthermore, fully-spectral rendering has received considerable attention in the computer graphics field in

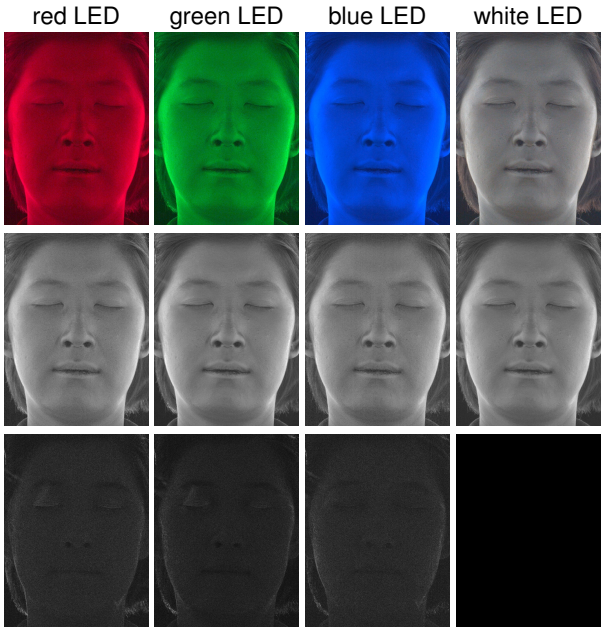
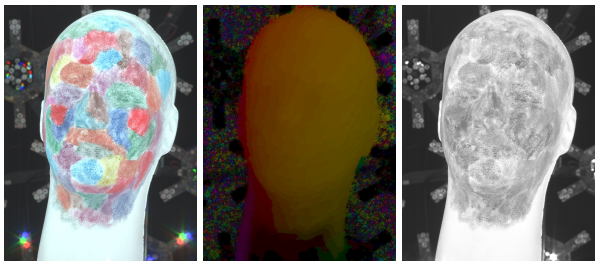


Figure 10: **Top row:** Ground truth polarization difference images, computed by photographing a subject under cross- and parallel-polarized lighting conditions for each spectrum. **Middle row:** Monochrome ground truth polarization difference images for each spectral channel, computed via RGB to grayscale conversion of the images of the top row. These are then scaled based on the calibrated LED intensities as observed by the camera. These images are qualitatively similar across spectral channels, indicating that the amount of light reflected specularly from the skin does not depend on the incident spectrum, justifying our technique of multispectral polarization promotion. **Bottom row:** Absolute difference across the images of the middle row. From left to right: |White - Red|, |White - Green|, |White - Blue|, and the trivially 0-valued |White-White|. Quantitative average absolute errors are $0.016, 0.0016, 0.0012$ for a large region of the subject's cheek, for red, green, and blue respectively as compared with the white LED. Errors are relative to incident light of intensity 1.0.



a. mannequin b. visualized flow c. white LED

Figure 11: **a:** A color image of the painted mannequin head used for the optical flow experiments. **b:** The flow field visualized by naïvely flowing one white LED image to another. Smoothly varying pixel colors indicate rigid motion. Hue represents the direction of the per-pixel motion vector, and saturation represents its magnitude. **c:** The image of the mannequin lit by the white LED, the target image for the flow calculations of Fig. 12.

recent years, and several commercial global illumination rendering software programs now include spectral information. It would be of theoretical interest to evaluate how our multispectral texture maps could be used in spectral global illumination rendering, particularly as the sub-surface scattering appearance for a given illuminant can be "baked" into our diffuse albedo, albeit for the full sphere of incident illumination.

Conclusion

In this work, we have demonstrated that high-resolution facial scanning with a light stage system can be efficiently achieved using monochrome cameras and colored LEDs, with advantages over the usual setup with color cameras and white LEDs. Only a few more images are required, equivalent to the number of added spectral channels, and the increased camera sensitivity can be used to shorten the exposure times to speed up the scan and/or reduce the light on the subject. In the case where only one spectral channel (e.g. white) is polarized, we introduced an approach to hallucinate cross- and parallel-polarized images for the remaining channels, as required for diffuse albedo texture map generation. We also introduced a novel multispectral optical flow technique based on complementary flow, enabling our multispectral 3D scanning technique to be applied to live subjects.

Acknowledgments

The authors wish to thank Randy Hill, Kathleen Haase, Hao Li, Christina Trejo, Etienne Danvoye, Chris Battson, and Matt Whalen 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

[1] B. Ajdin, M. Finckh, C. Fuchs, J. Hanika, and H. Lensch. Compressive higher-order sparse and low-rank acquisition with a hyperspectral light stage. Technical report, Univ. of Tuebingen, 2012.

[2] O. Alexander, G. Fyffe, J. Busch, X. Yu, R. Ichikari, A. Jones, P. Debevec, J. Jimenez, E. Danvoye, B. Antionazzi, et al. Digital ira: Creating a real-time photoreal digital actor. In *ACM SIGGRAPH 2013 Posters*, page 1. ACM, 2013.

[3] O. Alexander, M. Rogers, W. Lambeth, J.-Y. Chiang, W.-C. Ma, C.-C. Wang, and P. Debevec. The digital emily project: Achieving a photorealistic digital actor. *IEEE Computer Graphics and Applications*, 30(4):20–31, 2010.

[4] T. Beeler, B. Bickel, P. Beardsley, B. Sumner, and M. Gross. High-quality single-shot capture of facial geometry. In *ACM Transactions on Graphics (ToG)*, volume 29, page 40. ACM, 2010.

[5] D. Bradley, W. Heidrich, T. Popa, and A. Sheffer. High resolution passive facial performance capture. In *ACM transactions on graphics (TOG)*, volume 29, page 41. ACM, 2010.

[6] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.

[7] P. Debevec, T. Hawkins, C. Tchou, H.-P. Duiker, W. Sarokin, and M. Sagar. Acquiring the reflectance field of a human face. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 145–156. ACM Press/Addison-Wesley Publishing Co., 2000.

[8] P. Debevec, A. Wenger, C. Tchou, A. Gardner, J. Waese, and T. Hawkins. A lighting reproduction approach to live-action compositing. In *Proc. 29th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '02, pages 547–556, New York, NY, USA, 2002. ACM.

[9] A. Fenies, C. Kampouris, H. Deng, and A. Ghosh. Improving image-based lighting reproduction using a multispectral light stage. CVMP Short Papers, 2017.

[10] I. Fryc, S. W. Brown, and Y. Ohno. Spectral matching with an led-based spectrally tunable light source. In *Optics & Photonics 2005*, pages 59411I–59411I. International Society for Optics and Photonics, 2005.

[11] Y. Furukawa and J. Ponce. Dense 3d motion capture for human faces. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1674–1681. IEEE, 2009.

[12] G. Fyffe, P. Graham, B. Tunwattanapong, A. Ghosh, and P. Debevec. Near-instant capture of high-resolution facial geometry and reflectance. In *Computer Graphics Forum*, volume 35, pages 353–363. Wiley Online Library, 2016.

[13] G. Fyffe, A. Jones, O. Alexander, R. Ichikari, P. Graham, K. Nagano, J. Busch, and P. Debevec. Driving high-resolution facial blendshapes with video performance capture. In *ACM SIGGRAPH 2013 Talks*, SIGGRAPH '13, pages 33:1–33:1, New York, NY, USA, 2013. ACM.

[14] G. Fyffe, K. Nagano, L. Huynh, S. Saito, J. Busch, A. Jones, H. Li, and P. Debevec. Multi-view stereo on consistent face topology. In *Computer Graphics Forum*, volume 36, pages 295–309. Wiley Online Library, 2017.

[15] G. Fyffe, X. Yu, and P. Debevec. Single-shot photometric stereo by spectral multiplexing. In *Computational Photography (ICCP), 2011 IEEE International Conference on*, pages 1–6. IEEE, 2011.

[16] P. Garrido, L. Valgaerts, C. Wu, and C. Theobalt. Reconstructing detailed dynamic face geometry from monocular video. *ACM Trans. Graph.*, 32(6):158–1, 2013.

[17] A. Ghosh, G. Fyffe, B. Tunwattanapong, J. Busch, X. Yu, and P. Debevec. Multiview face capture using polarized spherical gradient illumination. *ACM Transactions on Graphics (TOG)*, 30(6):129, 2011.

[18] A. Ghosh, T. Hawkins, P. Peers, S. Frederiksen, and P. Debevec. Practical modeling and acquisition of layered facial reflectance. In *ACM Transactions on Graphics (TOG)*, volume 27, page 139. ACM, 2008.

[19] M. Goel, E. Whitmire, A. Mariakakis, T. S. Saponas, N. Joshi, D. Morris, B. Guenter, M. Gavriliu, G. Borriello, and S. N. Patel. Hypercam: hyperspectral imaging for ubiquitous computing applications. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 145–156. ACM, 2015.

[20] P. Graham, B. Tunwattanapong, J. Busch, X. Yu, A. Jones, P. Debevec, and A. Ghosh. Measurement-based synthesis of facial microgeometry. In *Computer Graphics Forum*, volume 32, pages 335–344. Wiley Online Library, 2013.

[21] J. Gu and C. Liu. Discriminative illumination: Per-pixel classification of raw materials based on optimal projections of spectral brdf. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 797–804, June 2012.

[22] P.-L. Hamon, J. Harmer, S. Penn, and N. Scapel. Gravity: Motion control and face integration. In *ACM SIGGRAPH 2014 Talks*, SIGGRAPH '14, pages 35:1–35:1, New York, NY, USA, 2014. ACM.

[23] C. Hernández, G. Vogiatzis, G. J. Brostow, B. Stenger, and R. Cipolla. Non-rigid photometric stereo with colored lights. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8. IEEE, 2007.

[24] K. Hirakawa and T. W. Parks. Adaptive homogeneity-directed demosaicing algorithm. *IEEE Transactions on Image Processing*, 14(3):360–369, 2005.

[25] H. Huang, J. Chai, X. Tong, and H.-T. Wu. Leveraging motion capture and 3d scanning for high-fidelity facial performance acquisition. In *ACM Transactions on Graphics (TOG)*, volume 30, page 74. ACM, 2011.

[26] H. W. Jensen, S. R. Marschner, M. Levoy, and P. Hanrahan. A practical model for subsurface light transport. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 511–518. ACM, 2001.

[27] A. Kimachi, H. Ikuta, Y. Fujiwara, M. Masumoto, and H. Matsuyama. Spectral matching imager using amplitude-modulation-coded multispectral light-emitting diode illumination. *Optical engineering*, 43(4):975–985, 2004.

[28] M. Kitahara, T. Okabe, C. Fuchs, and H. Lensch. Simultaneous estimation of spectral reflectance and normal from a small number of images. In *Proceedings of the 10th International Conference on Computer Vision Theory and Applications*, VISAPP 2015, pages 303–313, 2015.

[29] C. LeGendre, X. Yu, and P. Debevec. Efficient multispectral reflectance function capture for image-based relighting. In *Color and Imaging Conference*, volume 2016, pages 47–58. Society for Imaging Science and Technology, 2016.

[30] C. LeGendre, X. Yu, D. Liu, J. Busch, A. Jones, S. Pattanaik, and P. Debevec. Practical multispectral lighting reproduction. *ACM Transactions on Graphics (TOG)*, 35(4):32, 2016.

[31] J. P. Lewis, K. Anjyo, T. Rhee, M. Zhang, F. H. Pighin, and Z. Deng. Practice and theory of blendshape facial models. *Eurographics (State of the Art Reports)*, 1(8), 2014.

[32] W.-C. Ma, T. Hawkins, P. Peers, C.-F. Chabert, M. Weiss, and P. Debevec. Rapid acquisition of specular and diffuse normal maps from polarized spherical gradient illumination. In *Proceedings of the 18th Eurographics conference on Rendering Techniques*, pages 183–194. Eurographics Association, 2007.

[33] S. P. Mallick, T. E. Zickler, D. J. Kriegman, and P. N. Belhumeur. Beyond lambert: Reconstructing specular surfaces using color. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pages 619–626. Ieee, 2005.

[34] S. K. Nayar, X.-S. Fang, and T. Boult. Separation of reflection components using color and polarization. *International Journal of Computer Vision*, 21(3):163–186, 1997.

[35] J.-I. Park, M.-H. Lee, M. D. Grossberg, and S. K. Nayar. Multispectral imaging using multiplexed illumination. In *2007 IEEE 11th International Conference on Computer Vision*, pages 1–8. IEEE, 2007.

[36] R. Shrestha and J. Y. Hardeberg. Multispectral imaging using led illumination and an rgb camera. In *Color and Imaging Conference*, volume 2013, pages 8–13. Society for Imaging Science and Technology, 2013.

[37] B. Tunwattanapong, G. Fyffe, P. Graham, J. Busch, X. Yu, A. Ghosh, and P. Debevec. Acquiring reflectance and shape from continuous spherical harmonic illumination. *ACM Transactions on graphics (TOG)*, 32(4):109, 2013.

[38] A. Wenger, A. Gardner, C. Tchou, J. Unger, T. Hawkins, and P. Debevec. Performance relighting and reflectance transformation with time-multiplexed illumination. In *ACM Transactions on Graphics (TOG)*, volume 24, pages 756–764. ACM, 2005.

[39] A. Wenger, T. Hawkins, and P. Debevec. Optimizing color matching in a lighting reproduction system for complex subject and illuminant spectra. In *ACM International Conference Proceeding Series*, volume 44, pages 249–259, 2003.

[40] C. A. Wilson, A. Ghosh, P. Peers, J.-Y. Chiang, J. Busch, and P. Debevec. Temporal upsampling of performance geometry using photometric alignment. *ACM Transactions on Graphics (TOG)*, 29(2):17, 2010.

[41] L. B. Wolff. Material classification and separation of reflection components using polarization/radiometric information. *Proc. of Image Understanding Workshop*, pages 232–244, 1989.

[42] L. B. Wolff. Using polarization to separate reflection components. In *Computer Vision and Pattern Recognition, 1989. Proceedings CVPR'89., IEEE Computer Society Conference on*, pages 363–369. IEEE, 1989.

[43] R. J. Woodham. Photometric method for determining surface orientation from multiple images. *Optical engineering*, 19(1):191139–191139, 1980.

[44] C. Zach, T. Pock, and H. Bischof. A duality based approach for realtime tv-l 1 optical flow. In *Joint Pattern Recognition Symposium*, pages 214–223. Springer, 2007.

Author Biography

Chloe LeGendre earned a B.S. in Engineering in 2009 from the University of Pennsylvania and a M.S. in Computer Science from the Stevens Institute of Technology in 2015. From 2011 to 2015, she was an applications scientist in imaging and augmented reality for L'Oreal USA Research and Innovation. She is currently pursuing a Ph.D. in Computer Science at the University of Southern California's Institute for Creative Technologies (USC ICT), advised by Professor Paul Debevec.

Kalle (Karl) Bladin received a M.S. degree in Media Technology Engineering from Linköping University in 2016. He was one of the main developers of planetary visualization for the astronomical visualization software OpenSpace at the American Museum of Natural History and later at Visualization Center C, Norrköping. He is currently a research programmer at the USC ICT, focusing on deployable virtual human scanning, modeling, and rendering.

Bipin Kishore earned a B.E. in Electronics & Communication in 2015 from M.S. Ramaiah Institute of Technology, India, and a M.S. in

Electrical Engineering from USC in 2016. Since 2017 he has been working as an Electrical Engineer on the Light Stages and other related systems at the USC ICT.

Xinglei Ren received a B.S. in Electrical and Electronic Engineering in 2016 from the University of Shanghai for Science and Technology in China and a sole award to attend Coventry University in U.K. studying Electrical and Electronic Engineering in 2015. She earned an M.S. in Electrical Engineering from USC in 2018. She is currently working at the USC ICT as an Electrical Engineer.

Xueming Yu received a B.S. in Electrical Engineering from Shanghai Jiao Tong University in 2005 and an M.S. in Computer Science from USC in 2010. He developed 3D display technologies at Sony Corporation R&D and the electronics and mechanics for several Light Stage and surface reflectance scanning systems at the USC ICT. In 2016 he joined Google VR as a Hardware Engineer.

Paul Debevec received a Ph.D. in Computer Science from the University of California, Berkeley in 1996 under Professor Jitendra Malik specializing in photogrammetry and image-based rendering, with postdoctoral work in High Dynamic Range Imaging, Image-Based Lighting, and Reflectance Field capture. Light Stage systems from his laboratory at the USC Institute for Creative Technologies have been used to digitize models of actors in numerous feature films and were recognized with an Academy Scientific and Engineering Award in 2010. In 2016 he joined Google VR as a Senior Staff Engineer.

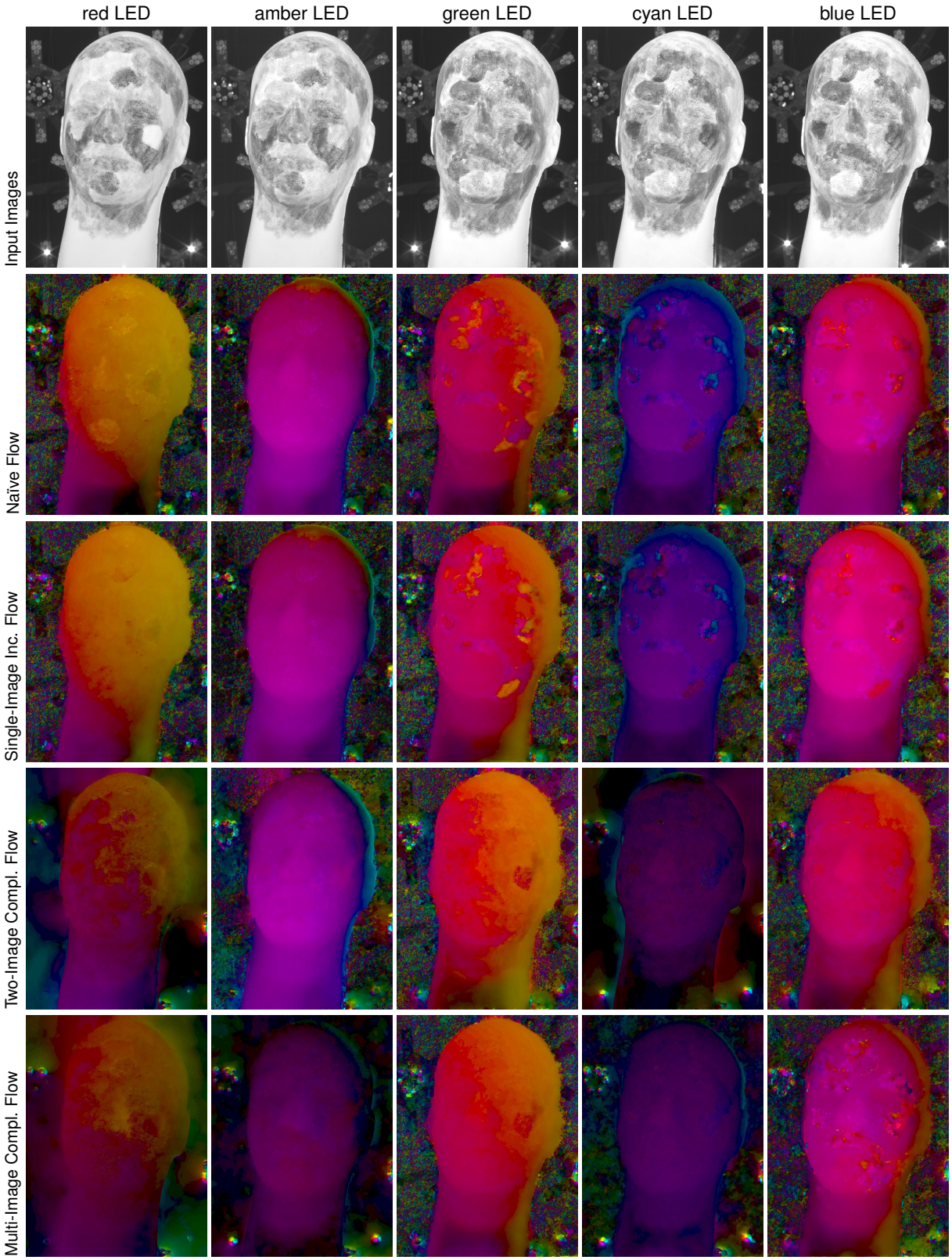


Figure 12: **Top row:** Input monochrome images of the painted mannequin head of Fig. 11a, as illuminated by each spectral channel. **Rows 2-5:** Visualizations of flow fields computed for each spectral channel, when aligning each image to the white LED image of Fig. 11c, using the four different flow techniques outlined the results section. Hue represents the direction of the per-pixel motion vector, while saturation represents the magnitude. Images have been scaled by 2.0 for display, increasing the apparent magnitude. The Two-Image Complementary Flow technique yields the smoothest flow fields representing the mannequin's rigid motion, best aligning the multispectral images for this challenging scan subject with diverse reflectance spectra.

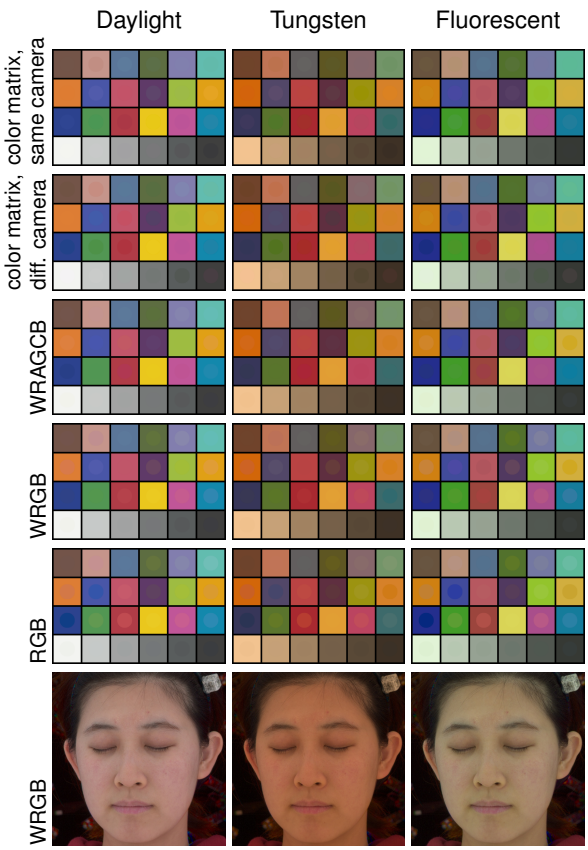


Figure 13: Matching color rendition with a 3×3 color matrix applied to color images, as compared to color channel mixing (Eq. 10) with multispectral monochrome imaging. Background squares are pixel values sampled from a color chart illuminated by three real world illuminants, photographed by a Canon 1DX camera. Foreground dots represent the closest achievable color rendition for each method. **Row 1:** 3×3 color matrix from an color image under white LED lighting, photographed by the same camera as the target, and **Row 2:** by a different camera from the target. **Rows 3, 4, and 5:** Monochrome imaging with six (WRAGCB), four (WRGB), and three (RGB) spectral channels. **Last row:** Scan subject of Fig. 7, cross-polarized images produced from a multispectral basis with Eq. 10. Color matrices alone cannot perfectly color-correct, especially when the input and target images are captured with different cameras. Quantitatively, multispectral imaging with monochrome cameras and six spectral channels improves color rendition for tungsten and daylight as compared with applying a color matrix alone. All charts have been converted to sRGB for display and are best viewed on a monitor.