

Single-Shot Photometric Stereo by Spectral Multiplexing

Graham Fyffe Xueming Yu Paul Debevec

Institute for Creative Technologies, University of Southern California
12015 Waterfront Drive, Los Angeles, CA

{fyffe, yu, debevec}@ict.usc.edu

Abstract

We propose a novel method for single-shot photometric stereo by spectral multiplexing. The output of our method is a simultaneous per-pixel estimate of the surface normal and full-color reflectance. Our method is well suited to materials with varying color and texture, requires no time-varying illumination, and no high-speed cameras. Being a single-shot method, it may be applied to dynamic scenes without any need for optical flow. Our key contributions are a generalization of three-color photometric stereo to more than three color channels, and the design of a practical six-color-channel system using off-the-shelf parts.

1. Introduction and Related Work

Photometric stereo is a powerful tool in the arsenal of 3D object acquisition techniques. Introduced by Woodham [11], photometric stereo methods estimate surface orientation (normals) by analyzing how a surface reflects light incident from multiple directions. Though it is not suitable for scenes with heavy occlusion or shadowing, photometric stereo has enjoyed widespread use and remains an active research topic. Nehab et al. [9] show that dense surface normal information may be used to improve the precision of scanned geometry. Ma et al. [8] show that photometric stereo captures fine-scale facial details missed by more direct geometric measurement techniques. Photometric stereo was originally applied to static scenes, but recent methods broaden its applicability to scenes with spatially varying color and dynamic motion. De Decker et al. [3] use multiple photographs together with spectral multiplexing to capture multiple illumination conditions with fewer photographs, for various applications including photometric stereo for scenes with motion and color variation. Kim et al. [5] improve on these results by explicitly modeling the effects of sensor crosstalk and changes in surface orientation due to subject motion. Still, *all* prior works that capture both color and normal rely on optical flow, and will fail if the scene contains enough motion or temporal inconsis-

tency. Our key motivation is to advance the state of the art in *single-shot* capture, because it has a fundamental advantage over multi-shot capture: it is trivially robust to any degree of motion or temporal inconsistency in the scene. Several prior works have achieved single-shot photometric stereo, by employing spectral multiplexing, but are all subject to certain limitations on the surface coloration. Woodham [11] proposed spectral multiplexing for three-source photometric stereo: using an off-the-shelf color camera, it is possible to image a scene as illuminated by three spectrally distinct light sources in a single photograph, and then use standard three-source photometric stereo methods to compute surface normals, provided the subject has uniform coloration. Klaudiny et al. [6] capture photometric surface normals for dynamic facial performances using spectrally multiplexed three-source illumination, but apply white makeup to the subject. Hernandez et al. [4] provide a detailed characterization of pixel intensities resulting from spectrally multiplexed illumination and describe a method for automatic calibration of the intrinsics of the apparatus. All of these previous works for single-shot photometric stereo assume that the materials in the scene have constant chromaticity, meaning that the spectral distribution of the surface reflectance varies only by a uniform scale factor. In this paper, we relax the constant chromaticity restriction, enabled by capturing images with a greater number of spectrally distinct color channels. We show that just a *single* multi-spectral photograph of a subject provides enough information to recover both the full-color reflectance and the surface normals on a per-pixel basis.

Other works have explored related aspects of multi-spectral capture. Wenger et al. [10] built a nine-channel light source made from different LEDs and filters for improved lighting reproduction on faces. Christensen et al. [2] show that multiple color channels provide useful information to photometric stereo, and note that more than three color channels may provide even further information. However, their work is in the context of using multiple photographs under differing illumination, and so it is not spectrally multiplexed in the sense that we are discussing here.

2. Spectrally Multiplexed Photometric Stereo

In spectrally multiplexed photometric stereo, a scene is illuminated by multiple spectrally distinct light sources, and photographed by a camera system configured to capture multiple spectrally distinct color channels. Following the notation of [4], the pixel intensity $c_k(x, y)$ for color channel k at pixel (x, y) is given by:

$$c_k(x, y) = \sum_j \mathbf{l}_j^T \mathbf{n}(x, y) \int E_j(\lambda) R(x, y, \lambda) S_k(\lambda) d\lambda, \quad (1)$$

where \mathbf{l}_j is the direction toward the j th light, λ represents wavelength, $E_j(\lambda)$ is the spectral distribution of the j th light (presumed distant), $\mathbf{n}(x, y)$ and $R(x, y, \lambda)$ are the surface normal and spectral distribution of the reflectance at pixel (x, y) (presumed Lambertian), and $S_k(\lambda)$ is the spectral response of the camera sensor for the k th color channel. Dropping the (x, y) indices, and considering discrete wavelengths instead of a continuous wavelength domain, we may write (1) in matrix form (with J lights, N discrete wavelengths, and K color channels) as:

$$\mathbf{c} = \mathbf{S} \text{diag}(\mathbf{r}) \mathbf{E} \mathbf{L} \mathbf{n}, \quad (2)$$

where $\mathbf{c} = [c_1, c_2, \dots, c_K]^T$, $\mathbf{S}(k, i) = S_k(\lambda_i)$, $\mathbf{r} = [R(x, y, \lambda_1), R(x, y, \lambda_2), \dots, R(x, y, \lambda_N)]^T$, $\mathbf{E}(i, j) = E_j(\lambda_i)$, and $\mathbf{L} = [\mathbf{l}_1, \mathbf{l}_2, \dots, \mathbf{l}_J]^T$. This is equivalent to the following system of bilinear equations in \mathbf{r} and \mathbf{n} :

$$c_k = \mathbf{r}^T \text{diag}(\mathbf{s}_k) \mathbf{E} \mathbf{L} \mathbf{n}, \quad k = 1 \dots K, \quad (3)$$

where $\mathbf{s}_k = [\mathbf{S}(k, 1), \mathbf{S}(k, 2), \dots, \mathbf{S}(k, N)]^T$. The system (3) is underdetermined, having $N + 3$ degrees of freedom but only K equations, and therefore requires $N + 3 - K$ additional constraints to regularize the system. In previous work with three color channels ($K = 3$), the required N constraints are given implicitly by restricting the reflectance to have constant chromaticity [11, 4]. Thus \mathbf{r} is presumed constant (with a scalar albedo factor absorbed into \mathbf{n}), reducing (3) to a system of linear equations. Our key contribution is to remove the requirement that \mathbf{r} be restricted to constant chromaticity. We *relax* this restriction, allowing \mathbf{r} to vary in some D -dimensional linear basis \mathbf{B} , yielding:

$$c_k = \hat{\mathbf{r}}^T \mathbf{B}^T \text{diag}(\mathbf{s}_k) \mathbf{E} \mathbf{L} \mathbf{n}, \quad k = 1 \dots K, \quad (4)$$

where $\hat{\mathbf{r}}$ is a D dimensional vector representing reflectance in the reduced basis. We may lump the reflectance basis, the sensor responses, and the illumination into a single matrix per color channel $\mathbf{M}_k = \mathbf{B}^T \text{diag}(\mathbf{s}_k) \mathbf{E} \mathbf{L}$, yielding:

$$c_k = \hat{\mathbf{r}}^T \mathbf{M}_k \mathbf{n}, \quad k = 1 \dots K. \quad (5)$$

We may then choose $D = K - 2$, so that the system (5) has $K + 1$ degrees of freedom and K equations, and we remove the final degree of freedom with $\|\mathbf{n}\| = 1$. We may

also increase robustness by lowering the dimensionality of the basis further ($D < K - 2$), causing (5) to be overdetermined. Least squares solutions to overdetermined systems of bilinear equations can be obtained using a normalized iterative algorithm [1], which naturally makes use of the normalization $\|\mathbf{n}\| = 1$. We make the modification that instead of requiring the first non-zero component of the normalized unknown to be positive, we require the z component of the surface normal (\mathbf{n}_z) to be positive, where z is defined to face toward the camera. The normalized iterative algorithm operates as follows:

```

 $\mathbf{n} \leftarrow [0, 0, 1]^T$ 
for several iterations do
   $\hat{\mathbf{r}} \leftarrow [\mathbf{M}_1 \mathbf{n}, \mathbf{M}_2 \mathbf{n}, \dots, \mathbf{M}_K \mathbf{n}]^T \setminus [c_1, c_2, \dots, c_K]^T$ 
   $\mathbf{n} \leftarrow [\mathbf{M}_1^T \hat{\mathbf{r}}, \mathbf{M}_2^T \hat{\mathbf{r}}, \dots, \mathbf{M}_K^T \hat{\mathbf{r}}]^T \setminus [c_1, c_2, \dots, c_K]^T$ 
   $\mathbf{n} \leftarrow \text{sign}(\mathbf{n}_z) \mathbf{n} / \|\mathbf{n}\|$ 
end for

```

where $\mathbf{A} \setminus \mathbf{b} = \arg \min_{\mathbf{x}} \|\mathbf{A} \mathbf{x} - \mathbf{b}\|^2$. One practical implication of our method is that capturing full-color RGB reflectance along with surface normals requires multi-spectral photography with at least *five* color channels. Intuitively this makes sense, as RGB reflectance has three degrees of freedom, and surface normals have two.

3. Apparatus

We realize our method with a six-color-channel implementation that enables simultaneous capture of surface normals and three-color-channel reflectance, overdetermined by one degree of freedom for robustness, though other configurations are possible. We illuminate a scene with three light sources, each a cluster of differently colored LED

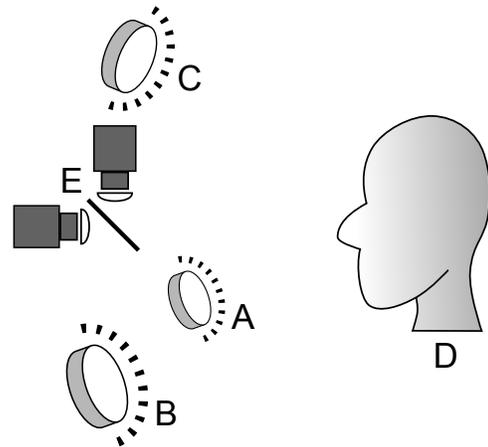


Figure 1. Schematic view of the apparatus. Three spectrally distinct light sources (A, B, C) illuminate a subject (D), who is recorded by a multi-spectral camera system (E), consisting of two ordinary cameras filtered by Dolby “left eye” and “right eye” dichroic filters, and aligned using a beam splitter.

lights with filters. We photograph the scene with a camera system configured to measure six different channels of the visible light spectrum. Figure 1 offers a schematic view of the apparatus.

To obtain six-channel photographs, we use a beam splitter to align two ordinary color cameras, and place a Dolby “left eye” dichroic filter over one camera lens, and a Dolby “right eye” dichroic filter over the other camera lens (as depicted in figure 1). Together, the two Dolby filters separate the visible spectrum into six non-overlapping bands, plotted in figure 2. We use Grasshopper cameras from Point Grey Research, which are easily synchronized to capture simultaneous photographs, and have a nearly linear intensity response curve. The cameras themselves are not modified, and standard color demosaicing algorithms may be used, since the data is captured as two separate three-channel photographs. A drawback of placing Dolby filters in front of the lenses is that more than half of the light reaching the sensors is lost. Nevertheless, we obtained well-exposed, low-noise images with the Grasshopper cameras in a real-time capture context. In applications requiring higher signal to noise ratios, three-chip color cameras could be employed, which have less inherent light loss than cameras based on Bayer color filter arrays.

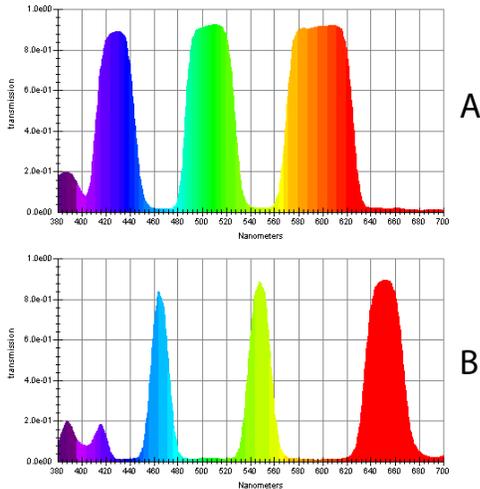


Figure 2. Spectral distribution plots of the Dolby “right eye” filter (A) and Dolby “left eye” filter (B).

The light sources are clusters of different combinations of violet, blue, cyan, green, orange and red LEDs. The spectral distributions of the violet, cyan and orange LEDs approximately coincide with the Dolby “right eye” filter, and the blue, green and red LEDs approximately coincide with the Dolby “left eye” filter. However, some of the LED colors overlap both the “left eye” and “right eye” Dolby filters, reducing any signal that may be encoded in the relationships between color channels. To eliminate this overlap, we also

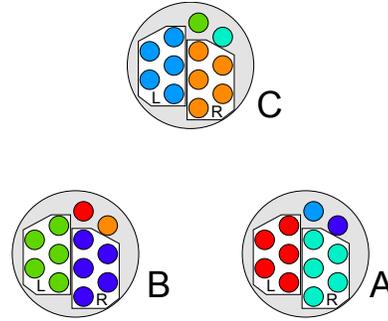


Figure 3. Schematic view of the three light sources (A, B, C). Six different colors of LEDs are arranged in clusters, some filtered by a Dolby “left eye” filter (L) or a Dolby “right eye” filter (R).

filter the LED clusters with Dolby “left eye” or “right eye” filters as appropriate. Figure 3 depicts the specific arrangement of colored LEDs used in the three light sources, which are as follows:

- A) 5 red LEDs with Dolby “left eye” filter, 5 cyan LEDs with Dolby “right eye” filter, 1 blue LED, 1 violet LED.
- B) 5 green LEDs with Dolby “left eye” filter, 5 violet LEDs with Dolby “right eye” filter, 1 red LED, 1 orange LED.
- C) 5 blue LEDs with Dolby “left eye” filter, 5 orange LEDs with Dolby “right eye” filter, 1 green LED, 1 cyan LED.

These light sources provide three distinct spectral distributions, plotted in figure 4. The LED clusters were chosen such that each light source has approximately equal brightness of red+orange, green+cyan and blue+violet light, appearing roughly white to the human eye, and having roughly equal overall intensity when viewed through either Dolby filter. Note that our method requires somewhat brighter and/or more LEDs than traditional three-light photometric stereo, due to the light loss from the filters. Finally, since our method presumes Lambertian reflectance, we place linear polarizing filters in front of the camera system and light sources, tuned to cancel out specular reflections on the subject. In common with prior work, these filters may be omitted in applications where the subject reflectance is predominantly diffuse. The beam splitter, LED lights, Dolby filters, and polarizing filters used in our apparatus are readily available and inexpensive, and any color cameras may be used so long as they can be synchronized to each other, making our system relatively easy to reproduce.

4. Calibration

The matrices M_k , $k = 1 \dots K$ in (5) required for our method may be obtained through the following calibration procedure. We photograph material samples with different

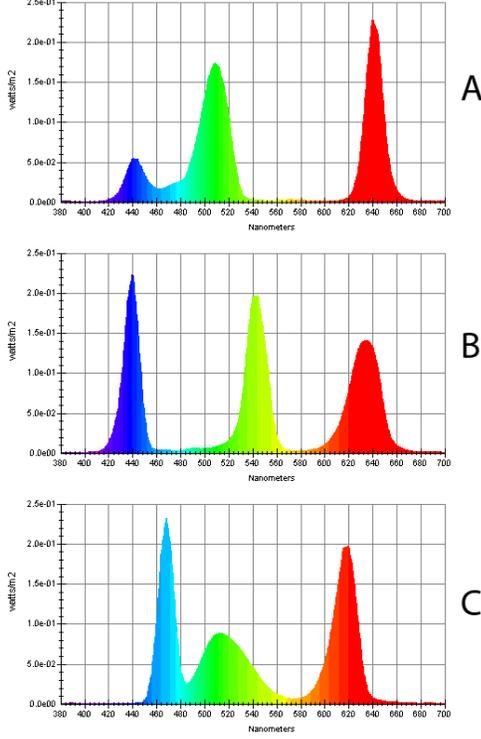


Figure 4. Spectral distribution plots of three LED light sources. A) red, cyan, dim blue, dim violet, B) green, violet, dim red, dim orange, C) blue, orange, dim green, dim cyan.

known reflectance values $\hat{\mathbf{r}}_t$ and surface normals \mathbf{n}_t , satisfying:

$$c_{k,t} = \hat{\mathbf{r}}_t^T \mathbf{M}_k \mathbf{n}_t, \quad k = 1 \dots K, \quad t = 1 \dots T, \quad (6)$$

where $c_{k,t}$ is the k th color channel of measurement t . We may then estimate \mathbf{M}_k , $k = 1 \dots K$ by:

$$\langle \mathbf{M}_k \rangle = \begin{bmatrix} \langle \hat{\mathbf{r}}_1 \mathbf{n}_1^T \rangle^T / \|\hat{\mathbf{r}}_1\|^\beta \\ \langle \hat{\mathbf{r}}_2 \mathbf{n}_2^T \rangle^T / \|\hat{\mathbf{r}}_2\|^\beta \\ \vdots \\ \langle \hat{\mathbf{r}}_T \mathbf{n}_T^T \rangle^T / \|\hat{\mathbf{r}}_T\|^\beta \end{bmatrix} \setminus \begin{bmatrix} c_{k,1} / \|\hat{\mathbf{r}}_1\|^\beta \\ c_{k,2} / \|\hat{\mathbf{r}}_2\|^\beta \\ \vdots \\ c_{k,T} / \|\hat{\mathbf{r}}_T\|^\beta \end{bmatrix}, \quad (7)$$

where $\langle \mathbf{A} \rangle$ is the lexicographic concatenation of the columns of \mathbf{A} , and $\beta \in (0 \dots 1)$ is a parameter to balance the importance of reflectance versus surface normal. In this work we use $\beta = \frac{1}{2}$. The choice of material samples used for calibration affects the accuracy of the method, since their spectral distributions effectively become a basis for recovered reflectance in a scene. Therefore, ideally the samples should be made from materials with similar spectral distributions as the materials in the scenes to be captured. If the materials in the scene are unknown in advance, an approximate calibration may be obtained with a set of generic materials. For the calibration in this

work, we use the twenty-four color swatches of a MacBeth ColorCheckerTM chart, photographed at five known orientations (frontal, up, down, left and right). We use linear sRGB color values for our reflectance basis, which are readily available for the color chart swatches. Figure 5 shows the color chart photographs used for calibration, and the reconstructed reflectance and surface normals after calibration. Using this many materials and orientations for calibration overconstrains the system in (7), resulting in some residual error. Figure 6 tabulates the per-swatch reconstruction error over the photographs used for calibration.

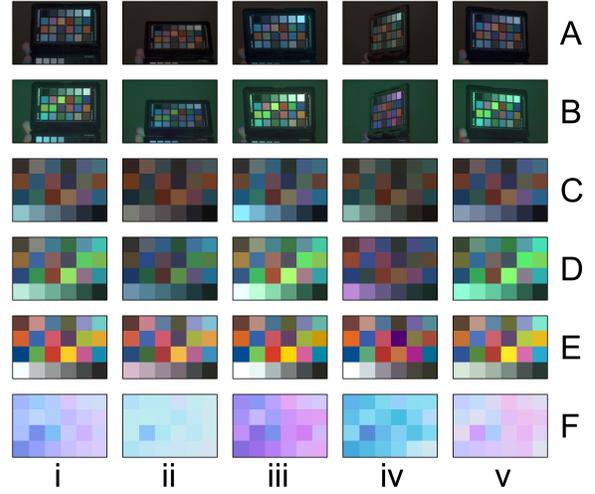


Figure 5. Calibration results. Columns are the different orientations of the color chart: i) frontal, ii) up, iii) down, iv) left, v) right. A) First three color channels of input photographs. B) Last three color channels of input photographs. C,D) A,B sampled at chart swatch centers, averaged over 5×5 pixel windows. E) Recovered reflectance. F) Recovered surface normals.

0.328 18.7°	0.097 21.3°	0.129 16.0°	0.340 27.4°	0.261 24.8°	0.365 28.3°
0.152 15.7°	0.179 20.8°	0.130 16.2°	0.212 9.54°	0.280 18.9°	0.293 30.4°
0.204 17.9°	0.562 41.2°	0.300 30.2°	0.269 15.9°	0.227 14.7°	0.324 20.6°
0.454 35.4°	0.273 22.4°	0.197 12.5°	0.203 15.6°	0.291 31.9°	0.375 41.6°

Figure 6. Reconstruction error after calibration, per swatch. Top number: relative RMSE of reflectance. Bottom number: RMSE of surface normal, in degrees. Overall relative RMSE of reflectance is 0.288, and overall RMSE of surface normal is 24.3°.

5. Results and Discussion

We show results for our proposed method using the described six-color-channel camera system, and the linear sRGB reflectance basis with generic color-chart-based calibration. Processing times are about one minute per frame on a single CPU using a straightforward implementation. Figure 7 shows results for the color chart, at novel orientations. The recovered reflectance appears stable, and the recovered surface normals extrapolate beyond the normals used in the calibration. However, the reconstruction errors tabulated in Figure 6 indicate that a generic calibration results in significant bias in the reconstruction. This is primarily caused by variation in the reflectance spectral distributions of the different materials, which cannot be represented exactly by a three-dimensional basis. This bias is also noticeable in results on a toy fish with highly saturated reflectance, shown in Figure 8. Figure 9 shows the input photographs and

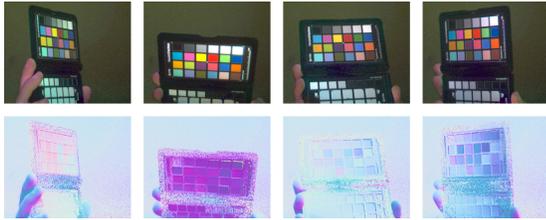


Figure 7. Results for the color chart, at novel orientations. Top: recovered reflectance. Bottom: recovered surface normals.



Figure 8. Recovered surface normals and reflectance of a toy fish.

reconstruction result for a human face, selected from a sequence in which the subject is talking and looking around. We also obtained a high-quality surface normal data set of the same subject in a similar pose using the method of Ma et al. [8] to serve as ground truth. We registered the ground truth data to one frame of our input data using optical flow, and drew a mask to separate the face from the background. (The eyes were also masked since they were closed in the ground truth data.) Figure 10 shows the RMSE of the recovered surface normals. As expected, the normals recovered using the generic calibration have significant bias. However, [4] obtains good normals for faces using a one-dimensional basis, so a higher-dimensional basis should be sufficient given a scene-dependent calibration. To obtain an

approximate scene-dependent calibration, we modeled the bias in the recovered normals with respect to the ground truth normals as a linear transform. We then applied the inverse transform to correct the normals, and finally fed the reflectance and corrected normals back into (7) to compute the scene-dependent calibration. (All of these operations considered only those pixels within the masked region.) We then used this scene-dependent calibration to process the entire face sequence. Figure 11 shows results for every tenth frame of the sequence. The color variation in the skin (freckles, lips, etc.) is captured without corrupting the surface normals. Since no special handling of shadows is done, shadow regions around prominent features such as the nose exhibit artifacts, which are most noticeable in the recovered surface normals. Generally, photometric stereo approaches must take shadowing or visibility into account, or else suffer from such artifacts wherever some of the lights are not visible to the surface.

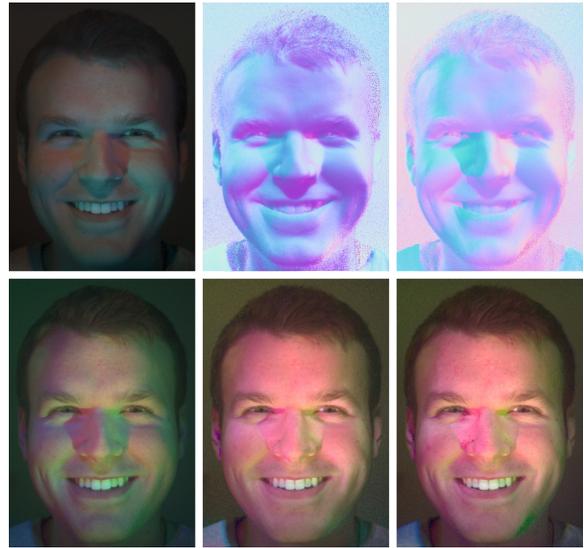


Figure 9. Results for a human face, selected from a dynamic sequence. Left top: first three color channels of input. Left bottom: last three color channels of input. Middle: recovered surface normals and reflectance, generic calibration. Right: recovered surface normals and reflectance, scene-dependent calibration.

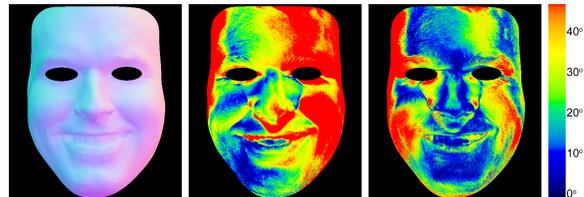


Figure 10. Comparison to ground truth. Left: ground truth surface normals using [8], registered using optical flow. Middle: surface normal error, generic calibration (38.3° RSME). Right: surface normal error, scene-dependent calibration (28.6° RSME).

6. Conclusion and Future Work

We have proposed a method to generalize spectrally multiplexed photometric stereo to more than three color channels, allowing, for the first time to our knowledge, true simultaneous capture of per-pixel photometric normals and full color reflectance. This enables new applications of photometric stereo, in dynamic scenes with spatially varying color and enough motion that optical flow methods fail. We implemented a practical six-color-channel apparatus using readily available parts. We showed results demonstrating that the method works for a variety of subjects, including a human face. In future work, we would like to address some of the limitations in our method. For scenes with few distinct materials, such as a human face, bias in the reconstruction caused by spectral variation is alleviated with scene-dependent calibration, but the scene-dependent calibration technique we employed is cumbersome. We would prefer to adapt the automatic scene-dependent calibration technique of [4]. For scenes with many distinct materials, bias may still persist even with scene-dependent calibration, and it may be necessary to increase the number of color channels captured by the camera system to combat this bias. Note that increasing the number of color channels from three to six already allows our system to handle more distinct materials in a scene than previous methods. Artifacts caused by shadows could possibly be addressed by adapting existing methods for handling shadows in photometric stereo, such as the visibility subspace method in [7]. Finally, using brighter, more distant light sources and/or more light sources would increase the usable scene volume, and may improve the signal to noise ratio in the results.

Acknowledgments

We wish to thank Joel Jurik for sitting as subject, Dolby Laboratories, Inc. for providing the dichroic filters, and Andrew Jones, Jay Busch, Bill Swartout, Randall Hill, and Randolph Hall for their support and assistance.

References

- [1] E. Bai and Y. Liu. On the least squares solutions of a system of bilinear equations. In *Proceedings of the 44th IEEE Conference on Decision and Control, and the European Control Conference*, pages 1197–1202. Seville, 2005. 2
- [2] P. H. Christensen and L. G. Shapiro. Three-dimensional shape from color photometric stereo. *International Journal of Computer Vision*, 13:213–227, 1994. 10.1007/BF01427152. 1
- [3] B. De Decker, J. Kautz, T. Mertens, and P. Bekaert. Capturing multiple illumination conditions using time and color multiplexing. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 2536–2543, 2009. 1



Figure 11. Recovered surface normal and reflectance for a human face, showing every tenth frame of a dynamic sequence.

- [4] C. Hernandez and G. Vogiatzis. Self-calibrating a real-time monocular 3d facial capture system. In *Proceedings International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT)*, 2010. 1, 2, 5, 6
- [5] H. Kim, B. Wilburn, and M. Ben-Ezra. Photometric stereo for dynamic surface orientations. In *Proceedings of the 11th European conference on Computer vision: Part I, ECCV'10*, pages 59–72, Berlin, Heidelberg, 2010. Springer-Verlag. 1
- [6] M. Kludiny, A. Hilton, and J. Edge. High-detail 3d capture of facial performance. In *3DPVT*, 2010. 1
- [7] N. P. Kostas Daniilidis, Petros Maragos, editor. *Visibility Subspaces: Uncalibrated Photometric Stereo with Shadows*, Heraklion, Greece, 2010. Springer-Verlag. 6
- [8] 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 *Rendering Techniques 2007: 18th Eurographics Symposium on Rendering*, pages 183–194, June 2007. 1, 5
- [9] D. Nehab, S. Rusinkiewicz, J. Davis, and R. Ramamoorthi. Efficiently Combining Positions and Normals for Precise 3D Geometry. *SIGGRAPH (ACM Transactions on Graphics)*, 24(3):536–543, 2005. 1
- [10] A. Wenger, T. Hawkins, and P. E. Debevec. Optimizing color matching in a lighting reproduction system for complex subject and illuminant spectra. In *Rendering Techniques*, pages 249–259, 2003. 1
- [11] R. J. Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19(1):139–144, 1980. 1, 2