Photometric Stereo for Dynamic Surface Orientations

Hyeongwoo Kim[1], Bennett Wilburn[2], and Moshe Ben-Ezra[2]

[1] KAIST, Daejeon, Republic of Korea
hyeongwoo.kim@kaist.ac.kr
[2] Microsoft Research Asia, Beijing, China
{bennett.wilburn,mosheb}@microsoft.com

Abstract. We present a photometric stereo method for non-rigid objects of unknown and spatially varying materials. The prior art uses time-multiplexed illumination but assumes constant surface normals across several frames, fundamentally limiting the accuracy of the estimated normals. We explicitly account for time-varying surface orientations, and show that for unknown Lambertian materials, five images are sufficient to recover surface orientation in one frame. Our optimized system implementation exploits the physical properties of typical cameras and LEDs to reduce the required number of images to just three, and also facilitates frame-to-frame image alignment using standard optical flow methods, despite varying illumination. We demonstrate the system's performance by computing surface orientations for several different moving, deforming objects.

1 Introduction

Photometric stereo [16] uses multiple images of an object illuminated from different directions to deduce a surface orientation at each pixel. In this work, we address accuracy limits for estimating surface orientations for dynamic scenes using photometric stereo, and in particular for deforming (non-rigid) objects. Photometric stereo for moving scenes is complicated because the world has only one illumination condition at a time, and the scene may move as one tries to change the lighting. Using a color camera and colored lights from different directions, one can measure shading for light from three directions in one image, but this only determines the surface orientation if the object reflectance is known and uniform.

The prior art for photometric stereo with deforming objects of varying or unknown materials uses time-multiplexed illumination (TMI) [15] to capture video while changing the lighting from frame to frame. Subsequent frames are aligned using optical flow, and the surface orientation is assumed to be constant across those frames. Assuming fixed surface normals for dynamic scenes is a contradiction and represents a fundamental accuracy limit for current TMI photometric stereo methods for non-rigid objects. For commonly occurring motions, we show this leads to significant errors in estimated surface orientations and albedos.

We present a photometric stereo method for deforming objects that is robust to changing surface orientations. We use time and color illumination multiplexing with three colors, but ensure an instantaneous measurement in every frame, either of the surface normal or of a subset of the material imaging properties. We use optical flow to account for varying motion at each pixel. Unlike the prior art, given accurate optical flow, our estimated surface normals are not corrupted if those normals are time-varying. Optical flow for TMI video is challenging because the intensity constancy assumption does not hold. Our optimized system implementation ensures constant illumination in one color channel, facilitating optical flow between subsequent frames using standard methods, despite varying illumination. Photometric stereo results for several deforming objects verify the performance of the system.

2 Background

Although the literature on shape capture of deforming objects is vast, Nehab et al. [9] observed that orientation-sensing technologies like photometric stereo are more accurate for high frequency shape details, while range sensing technologies (such as multi-view stereo) are better for low frequency shape. They devised an efficient method to combine the two forms of data to estimate precise geometry. These two forms of shape estimation are fundamentally different, so we will restrict our review to photometric stereo methods. The traditional photometric stereo [16] formulation assumes a static object imaged by a fixed camera under varying illumination directions. For a moving rigid object, many methods combine shading information with motion or multi-view stereo, assuming either fixed illumination (for example, [11, 1, 8]) or even varying lighting [6]. In this work, however, we aim to measure the surface orientation of deforming (non-rigid) objects, whose shape may vary from frame to frame, and whose motion cannot be represented simply as a rigid transformation.

Petrov [10] first addressed photometric stereo with multi-spectral illumination. One challenge of multi-spectral photometric stereo is the camera color measurements depend not only on the surface normal and light direction, but also on the interaction between the light spectra, material spectral responses, and the camera color spectral sensitivities. The method of Kontsevich et al. calibrates these dependencies using the image of the object itself, assuming the surface has a sufficient distribution of orientations [7]. The technique works for uncalibrated objects and materials, but is sensitive to the object geometry and unwieldy for multi-colored objects. Hernández et al. [5] presented a photometric stereo method that uses colored lights to measure surface normals on deforming objects. They show impressive results capturing time-varying clothing geometry, but the method requires the objects to consist of a single uniform material.

Wenger et al. [15] propose using time-multiplexed illumination (TMI) for photometric stereo. Their system uses high-speed video of an actor under 156 different lighting conditions and aligns images to target output frames using optical flow. Their goal is performance relighting, but they also compute surface normals and albedos for deforming objects and changing materials. Vlasic et al. [13] extend this idea to multi-view photometric stereo, using an array of cameras and time-multiplexed basis lighting. Both methods assume fixed normals across the images used to compute each output frame. Weise et al. [14] explicitly handle deforming objects with TMI to sense depth (not orientation) using a stereo phase-shift structured light technique.

De Decker et al. [3] combine time and color multiplexing to capture more illumination conditions in fewer frames than TMI alone. Their photometric stereo method does not explicitly address changing surface orientations. It also neglects light–sensor crosstalk, causing significant errors for common cameras (including theirs). The method computes optical flow using a filter that "removes the lighting, but preserves the texture" by normalizing for local brightness and contrast. For photometric stereo, however, the image texture and lighting are not separable. Imagine the dimples on a golf ball lit from one side, and then the other—the changing texture is itself the shading information. Assuming it to be a fixed feature for optical flow will corrupt the estimated normals.

In this paper, we describe how to use time and color multiplexing for photometric stereo given changing surface orientations. We start by adding a changing surface normal to the traditional photometric stereo formulation.

3 Dynamic Photometric Stereo with Time and Color Multiplexed Illumination

The observed intensity of a Lambertian surface with surface normal $\hat{\mathbf{n}}$, illuminated from direction $\hat{\mathbf{l}}$ is

$$I = \hat{\mathbf{l}} \cdot \hat{\mathbf{n}} \int S(\lambda)\rho(\lambda)\nu(\lambda)d\lambda, \tag{1}$$

where $S(\lambda)$ is the light energy distribution versus wavelength, $\rho(\lambda)$ is the material spectral reflectance, and $\nu(\lambda)$ is the camera spectral sensitivity. For fixed material, camera and light spectra, the integral is represented by the albedo, α:

$$I = \alpha\hat{\mathbf{l}} \cdot \hat{\mathbf{n}}. \tag{2}$$

If the surface is fixed, a minimum of three measurements with non-planar, known lighting directions are required to determine the normal and albedo[16]:

$$\begin{bmatrix} I_1 \\ I_2 \\ I_3 \end{bmatrix} = \alpha \begin{bmatrix} \hat{\mathbf{l}}_1 \\ \hat{\mathbf{l}}_2 \\ \hat{\mathbf{l}}_3 \end{bmatrix} \hat{\mathbf{n}} \tag{3}$$

For a dynamic scene, we assume the material reflectance is constant, but the surface normal varies between measurements. The system is now under-constrained:

$$\begin{bmatrix} I_1 \\ I_2 \\ I_3 \end{bmatrix} = \alpha \begin{bmatrix} \hat{\mathbf{l}}_1 \cdot \hat{\mathbf{n}}_1 \\ \hat{\mathbf{l}}_2 \cdot \hat{\mathbf{n}}_2 \\ \hat{\mathbf{l}}_3 \cdot \hat{\mathbf{n}}_3 \end{bmatrix} \tag{4}$$

Using a trichromatic camera and three lights of different colors, we measure shading under three different lighting directions simultaneously and thus for a single consistent surface orientation. Consider a camera with three color [3] channels labeled r, g, and b. Let us assume that each light, indexed by j, is from direction $\hat{\mathbf{l}}_\mathbf{j}$, and that each light j is composed of a weighted combination of a small number of light colors, indexed by k. For simplicity, first consider a single light of a single color k and intensity w_{kj}, and direction $\hat{\mathbf{l}}_\mathbf{j}$. The pixel intensity of a material illuminated by that light is

$$\mathbf{I} = \begin{bmatrix} I_r \\ I_g \\ I_b \end{bmatrix} = \begin{bmatrix} \alpha_{kr} \\ \alpha_{kg} \\ \alpha_{kb} \end{bmatrix} w_{kj} \hat{\mathbf{l}}_\mathbf{j}^\top \hat{\mathbf{n}} \tag{5}$$

Here, $(\alpha_{kr}, \alpha_{kg}, \alpha_{kb})^\top$ are the responses of each camera color channel to the material illuminated (from the normal direction) by light of color k. For example,

$$\alpha_{kr} = \int S_k(\lambda)\rho(\lambda)\nu_r(\lambda) d\lambda. \tag{6}$$

We refer to $\alpha_\mathbf{k} = (\alpha_{kr}, \alpha_{kg}, \alpha_{kb})^\top$ as a vector of "imaging coefficients." They are not just a property of a specific material; rather, they vary for each different combination of light, material and sensor colors. For a single light of direction $\hat{\mathbf{l}}_\mathbf{j}$ comprised of a linear combination of colors k, the measured pixel intensity is

$$\mathbf{I} = \begin{bmatrix} I_r \\ I_g \\ I_b \end{bmatrix} = \left(\sum_k \alpha_\mathbf{k} w_{kj} \right) \hat{\mathbf{l}}_\mathbf{j}^\top \hat{\mathbf{n}}. \tag{7}$$

For multiple lights, barring occlusions, we sum intensities due to each light:

$$\mathbf{I} = \begin{bmatrix} I_r \\ I_g \\ I_b \end{bmatrix} = \sum_j \left(\sum_k \alpha_\mathbf{k} w_{kj} \hat{\mathbf{l}}_\mathbf{j}^\top \right) \hat{\mathbf{n}} = \sum_k \left(\alpha_\mathbf{k} \left(\sum_j w_{kj} \hat{\mathbf{l}}_\mathbf{j}^\top \right) \right) \hat{\mathbf{n}} = \sum_k \left(\alpha_\mathbf{k} \mathbf{l_k}^\top \right) \hat{\mathbf{n}}, \tag{8}$$

where

$$\mathbf{l_k} = \sum_j w_{kj} \hat{\mathbf{l}}_\mathbf{j}. \tag{9}$$

Here, $\mathbf{l_k}$ can be considered the effective direction and intensity of light of color k. If $\alpha_\mathbf{k}$ are known and linearly independent, and $\mathbf{l_k}$ are known and linearly independent, then we can measure $\hat{\mathbf{n}}$ in a single image.

Of course, although the $\mathbf{l_k}$ may be known in advance for calibrated lights, the reflectance coefficients $\alpha_\mathbf{k}$ for materials in a dynamic scene are generally

[3] Color scientists might cringe at our usage of the words color, red, green, and blue for non-perceptual quantities. For the sake of readability, we use red, green and blue as a shorthand for visible spectra with most of the energy concentrated in longer, medium or shorter wavelengths, respectively. When we say the color of two lights are the same, we mean the spectra are identical up to a scale factor.

unknown. For scenes with spatially varying or unknown materials (and thus unknown α_k), we use additional time-multiplexed measurements with changing $\mathbf{l_k}$, producing a series of measurements:

$$\begin{aligned}
\mathbf{I}^t &= (\sum_k \alpha_{\mathbf{k}} \mathbf{l}_k^{t\top})\hat{\mathbf{n}}^t \\
\mathbf{I}^{(t+1)} &= (\sum_k \alpha_{\mathbf{k}} \mathbf{l}_k^{(t+1)\top})\hat{\mathbf{n}}^{(t+1)} \\
\mathbf{I}^{(t+2)} &= (\sum_k \alpha_{\mathbf{k}} \mathbf{l}_k^{(t+2)\top})\hat{\mathbf{n}}^{(t+2)} \\
&\vdots
\end{aligned} \tag{10}$$

We assume, for now, that we can align these measurements perfectly using optical flow. The $\mathbf{l}_k$ are known in advance, but the $\alpha_{\mathbf{k}}$ and normals are not. In general, Equation 10 is difficult to solve. If we use F frames and three light colors ($k = 3$), we have $9 + 2F$ unknowns for the reflectance coefficients and per-frame normals, but only $3F$ measurements. We need not, however, recover the surface normal for every frame. Instead, we will use one image with three spatially separated colored lights to measure the surface normal instantaneously, and use additional frames with carefully chosen lighting conditions to recover imaging coefficients α_k independently of the changing surface orientation.

We consider the minimum of three light colors; using more only adds more unknown imaging coefficients. With three sensor colors and three light colors, Equation 8 can be rewritten as

$$\mathbf{I} = \begin{bmatrix} I_r \\ I_g \\ I_b \end{bmatrix} = \sum_{k=1}^{3} (\alpha_{\mathbf{k}} \mathbf{l_k})\hat{\mathbf{n}} = \begin{bmatrix} \alpha_{1r} & \alpha_{2r} & \alpha_{3r} \\ \alpha_{1g} & \alpha_{2g} & \alpha_{3g} \\ \alpha_{1b} & \alpha_{2b} & \alpha_{3b} \end{bmatrix} \begin{bmatrix} \mathbf{l}_1^\top \\ \mathbf{l}_2^\top \\ \mathbf{l}_3^\top \end{bmatrix} \hat{\mathbf{n}}. \tag{11}$$

Now we will show that using four images, we can compute the unknown $\alpha_{\mathbf{k}}$ up to a single global scale factor. We capture three images, each lit by a single color, with the lighting directions being linearly independent and the colors being different for all images. For a point on the moving surface, this yields color pixel intensities $\mathbf{I_1}$, $\mathbf{I_2}$, and $\mathbf{I_3}$. The image for $\mathbf{I_1}$ is taken under illumination of color $k = 1$ with scaled direction $\mathbf{l_1} = w_{k1}\hat{\mathbf{l}}_\mathbf{1}$, and so on. We take another image using lights of all three colors from a single direction, producing the follow system:

$$\mathbf{I_1} = \alpha_\mathbf{1} \mathbf{l}_\mathbf{1}^\top \hat{\mathbf{n}}_\mathbf{1} = \alpha_\mathbf{1} s_1 \tag{12}$$

$$\mathbf{I_2} = \alpha_\mathbf{2} \mathbf{l}_\mathbf{2}^\top \hat{\mathbf{n}}_\mathbf{2} = \alpha_\mathbf{2} s_2 \tag{13}$$

$$\mathbf{I_3} = \alpha_\mathbf{3} \mathbf{l}_\mathbf{3}^\top \hat{\mathbf{n}}_\mathbf{3} = \alpha_\mathbf{3} s_3 \tag{14}$$

$$\mathbf{I_4} = (\alpha_\mathbf{1} + \alpha_\mathbf{2} + \alpha_\mathbf{3}) \mathbf{l}_\mathbf{4}^\top \hat{\mathbf{n}}_\mathbf{4} = (\alpha_\mathbf{1} + \alpha_\mathbf{2} + \alpha_\mathbf{3}) s_4 \tag{15}$$

We have used s_1 to represent the unknown scale factor $\mathbf{l_1}^\top \hat{\mathbf{n}}_\mathbf{1}$ in the first equation, and so on. Solving the top three equations for $\alpha_\mathbf{1}$, $\alpha_\mathbf{2}$, and $\alpha_\mathbf{3}$, respectively, and substituting into the fourth yields

$$\mathbf{I_4} = (\mathbf{I_1}/s_1 + \mathbf{I_2}/s_2 + \mathbf{I_3}/s_3)s_4, \tag{16}$$

or

$$\mathbf{I_4} = \begin{bmatrix}\mathbf{I}_1 \ \mathbf{I}_2 \ \mathbf{I}_3\end{bmatrix} \begin{bmatrix} s_4/s_1 \\ s_4/s_2 \\ s_4/s_3 \end{bmatrix} \tag{17}$$

We solve this system for s_1, s_2, and s_3 up to a scale factor $1/s_4$, and use Equations 12-14 to get $\alpha_{\mathbf{1}}$, $\alpha_{\mathbf{2}}$, and $\alpha_{\mathbf{3}}$ up to the same factor. As Equation 8 shows, this ambiguity does not prevent recovery of the normal using a fifth image taken with spatially separated colored lights. In practice, five different images may be too many to capture at video rates, and aligning the set of images may be difficult due to occlusions and varying illumination. In the next section, we explore ways to optimize this method.

4 Implementation

We have presented a theory for instantaneously measuring either surface orientation or imaging properties that vary with the materials. In this section, we investigate using fewer images, facilitating accurate optical flow to align those images, and using commonly available hardware.

Camera and Light Spectral Characteristics. A straightforward way to reduce the unknowns at each pixel, and thus require fewer images, is to ensure that some of the imaging coefficients are zero. We might try to use red, green and blue lights such that there is no "crosstalk" between lights and camera color sensors of different colors. Materials illuminated by only the green light, for example, would not register on the camera's red or blue color channels. This corresponds to the simplified component-wise (R, G, B) imaging model often used in computer graphics and also by a recent work on dynamic photometric stereo using colored lights [3]. Each material and light color is described by an RGB triplet, and the reflected intensity from a Lambertian surface with normal $\hat{\mathbf{n}}$ and reflectance $A = (A_R, A_G, A_B)$ lit by light of color $L = (L_R, L_G, L_B)$ from direction $\hat{\mathbf{n}}$ is

$$C = (C_R, C_G, C_B) = (A_R L_R, A_G L_G, A_B L_B)(\hat{\mathbf{n}} \cdot \hat{\mathbf{l}}). \tag{18}$$

We explored this approach using a typical single-chip color video camera, the Point Grey Research Flea2 FL2-08S2. With gamma correction off, the Flea2 has a linear response over most of its range. For lighting we use red, green and blue Luxeon K2 light emitting diodes (LEDs). These LEDs are inexpensive, bright, switch quickly (important for TMI), and have relatively narrow spectral power distributions.

Figure 1 shows the spectral characteristics of our camera and LEDs, and reveals two relevant properties. First, the spectra of the blue and green LEDs and the blue and green color sensors significantly overlap. Generally speaking, single-chip color sensors (as well as our own eyes) use color sensors with wide spectral responses for increased sensitivity, so crosstalk is unavoidable (in this case, more than one color sensor responds to the same light color). On the other

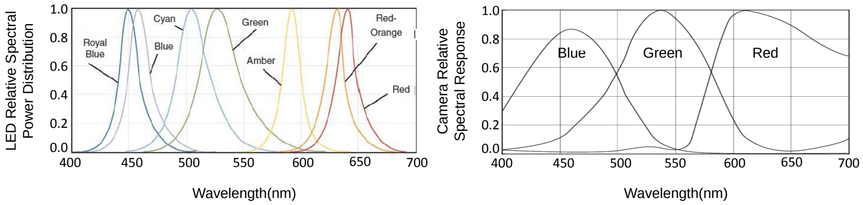

Fig. 1. Overlapping LED spectra and camera color responses necessitate using a complete imaging model, not a simplified component-wise RGB one. (Left) Relative spectral power distributions for different color Luxeon K2 LEDs. (Right) Relative spectral response for the red, green and blue pixels on the SONY ICX204 image sensor used in our camera.

hand, the red and blue color channels are decoupled; the red LED spectra has virtually no overlap with the blue color sensor, and vice versa. Because we are now assigning color labels like ”red” to our lights, we will switch to using capital letters instead of numbers to label light colors. We will use upper case R, G, and B for light colors, while still using lower case r, g, and b to for camera color channels. For the decoupled blue and red color channels in our system, we expect $\alpha_{Rb} = \alpha_{Br} = 0$. Using images of the patches on a Gretag Macbeth color checker [4] illuminated one color LED at a time, we verified that α_{Br} and α_{Rb} are negligible for our hardware. Unfortunately, the crosstalk for the red/green and green/blue color combinations is significant and varies greatly for different materials. We found that α_{Bg}/α_{Bb} varies across materials from 0.24 to 0.42, α_{Gb}/α_{Gg} varies from 0.08 to 0.29, α_{Rg}/α_{Rr} varies from 0.03 to 0.06, and α_{Gr}/α_{Gg} is on the order of a percent. These ratios are significant and change greatly from patch to patch, meaning that all non-zero imaging coefficients must be measured for any unknown material.

The LED and camera characteristics and the Macbeth experiment suggest an efficient way to eliminate two more imaging coefficients. We place Edmund Optics Techspec 550nm shortpass filters over the green LEDs to block the longer wavelengths sensed by the red camera color sensor, and Thorlabs FB650-40 filters over the red LEDs to ensure that they do not excite the green camera sensor. Now each measured color pixel corresponds to a much simpler set of equations. For an image taken with illumination from three spatially separated colored lights whose intensities and directions are described by $\mathbf{l_R}$, $\mathbf{l_G}$, and $\mathbf{l_B}$, Equation 8 yields

$$\begin{bmatrix} I_r \\ I_g \\ I_b \end{bmatrix} = \begin{bmatrix} \alpha_{Rr} & 0 & 0 \\ 0 & \alpha_{Gg} & \alpha_{Bg} \\ 0 & \alpha_{Gb} & \alpha_{Bb} \end{bmatrix} \begin{bmatrix} \mathbf{l}_R^\top \\ \mathbf{l}_G^\top \\ \mathbf{l}_B^\top \end{bmatrix} \hat{\mathbf{n}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \alpha'_{Gg} & \alpha'_{Bg} \\ 0 & \alpha'_{Gb} & \alpha'_{Bb} \end{bmatrix} \begin{bmatrix} \mathbf{l}_R^\top \\ \mathbf{l}_G^\top \\ \mathbf{l}_B^\top \end{bmatrix} (\alpha_{Rr}\hat{\mathbf{n}}) \qquad (19)$$

Here, we have substituted $\alpha'_{Gg} = \alpha_{Gg}/\alpha_{Rr}$, $\alpha'_{Bg} = \alpha_{Bg}/\alpha_{Rr}$, and so on.

We can compute the normal direction from Equation 19 if we know α'_{Gg}, α'_{Gg}, α'_{Gg}, and α'_{Gg}. These values can be measured using just two additional images: one with red and green lights from the same direction $\mathbf{l}_{RG}$, the other with red and blue lights from the same direction $\mathbf{l}_{RB}$. The first images gives

$$\begin{bmatrix} I_r \\ I_g \\ I_b \end{bmatrix} = \begin{bmatrix} \alpha_{Rr} & 0 & 0 \\ 0 & \alpha_{Gg} & \alpha_{Bg} \\ 0 & \alpha_{Gb} & \alpha_{Bb} \end{bmatrix} \begin{bmatrix} \mathbf{l}_{RG}^{\top} \\ \mathbf{l}_{RG}^{\top} \\ \mathbf{0}^{\top} \end{bmatrix} (\hat{\mathbf{n}}_{RG}) = \begin{bmatrix} \alpha_{Rr} \\ \alpha_{Gg} \\ \alpha_{Gb} \end{bmatrix} (\mathbf{l}_{RG}^{\top} \hat{\mathbf{n}}_{RG}). \tag{20}$$

Despite the unknown normal $\hat{\mathbf{n}}_{RG}$, we can solve for $\alpha'_{Gg} = \alpha_{Gg}/\alpha_{Rr} = I_r/I_g$ and $\alpha'_{Gb} = \alpha_{Gb}/\alpha_{Rr} = I_r/I_b$. Similarly, the second image with red and blue lights from direction $\mathbf{l}_{RB}$ determines α'_{Bg} and α'_{Bb}.

Time-Multiplexed Illumination and Optical Flow. Our system uses optical flow to align the two frames for measuring imaging coefficients to the frame illuminated with spatially separated red, green and blue lights. The red light is used for every frame. To facilitate optical flow, we set the red lighting to be constant and from the direction of the camera. Although the green and blue lights vary, they do not affect the red camera sensor, so the red video channel appears to have constant illumination. Setting the red light to arrive from the same direction as the camera prevents any shadows in the red channel of the video. We can robustly estimate optical flow for the red channel between adjacent frames using standard algorithms.

We output orientation measurements at half the video camera frame rate using the following lighting sequence:

$$R_c + G_c \big| R_c + G + B \big| R_c + B_c \big| R_c + G + B \ldots$$

Here, R_c, G_c, B_c, indicate red, green and blue lights from the direction of the camera, and G and B indicate the additional green and blue light directions used to estimate the normal. Each $R_c + G + B$ image is adjacent to an $R_c + G_c$ and an $R_c + B_c$ image.

Because our method measures material properties independently of the surface normal, the optical flow need not be pixel-accurate. As long as the alignment maps regions of the same material to each other, the surface normal estimate will be correct. Segmentation-based optical flow methods, for example, often have this property, even if subtle changes in shading from frame to frame may distort flow estimates within segments of the same material.

Hardware Design. Figure 2 shows a schematic of our system and the actual hardware. The setup has three spatially separated red, green and blue lights, labeled G, B, and R_c. The LEDs are positioned and filtered as described in the previous section. A simple microcontroller circuit triggers the LEDs and camera. We trigger the camera at 30Hz, but compute normal information for a video at half that rate. This is not a fundamental limit of our technique; upgrading to a 60Hz camera would enable normal map computations for a 30Hz sequence. Similar to Hernández et al., we use images of a diffuse plane at multiple known orientations to estimate the light intensities and directions $\mathbf{l}_G$, $\hat{\mathbf{l}}_{Rc}$, $\mathbf{l}_B$, $\hat{\mathbf{l}}_{Gc}$, and $\hat{\mathbf{l}}_{Gc}$. The lights next to the camera are assumed to have unit intensity, and the magnitudes of vectors $\mathbf{l}_G$ and $\mathbf{l}_B$ specify the intensity ratios between lights G

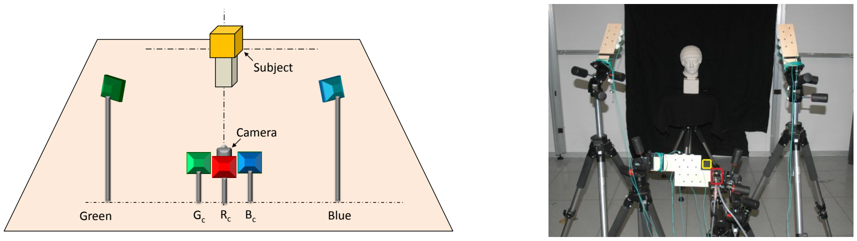

Fig. 2. A schematic diagram and the actual hardware used in our system. Each light is made of three LEDs with lenses mounted in a triangle pattern on the wooden boards (some LEDs shown on the boards are not used). The camera (outlined in red) aims through the square notch (shown in yellow) in the top right corner of the bottommost board.

and G_c, and B and B_c, respectively. For each set of three images, we use the optical flow method of Black and Anandan[2] (but using only the red channel) to compute the location of each point in the R_c+G+B image in the neighboring R_c+Gc and R_c+B_c images. The material imaging coefficients from those points are used to estimate the normals for the R_c+G+B frame.

5 Results

In this section, we present simulations to show the errors caused by (1) assuming constant normals for photometric stereo using alternating white lights, and (2) using a component-wise RGB imaging model in the presence of crosstalk. After that we show surface reconstructions and renderings produced using our method for challenging scenes.

5.1 Simulations

Changing Surface Orientations. Our first simulation investigates the accuracy of photometric stereo for Lambertian deforming objects using traditional TMI with alternating white lights. We will assume perfect optical flow to align the moving images, so the errors are due purely to the changing surface orientation between measurements. We simulated a system with three alternating white lights, capturing a rotating white surface with albedo $\alpha = 1.0$. The three measurements are the dot product of the normal and lighting directions: $I_1 = \hat{\mathbf{l}}_1 \cdot \hat{\mathbf{n}}_1$, $I_2 = \hat{\mathbf{l}}_2 \cdot \hat{\mathbf{n}}_2$, and $I_3 = \hat{\mathbf{l}}_3 \cdot \hat{\mathbf{n}}_3$. Combining these observations and assuming a constant normal is equivalent to solving the system $\mathbf{I} = \mathbf{L}\hat{\mathbf{n}_{\mathbf{c}}}$, where the rows of $\mathbf{L}$ are $\hat{\mathbf{l}}_1$, $\hat{\mathbf{l}}_2$, and $\hat{\mathbf{l}}_3$; and $\mathbf{I} = (I_1, I_2, I_3)^\top$.

We simulated a 30fps camera pointing along the negative z axis, viewing a surface at the origin with 1Hz rotational motion. 1Hz is actually a conservative number; people often turn their heads, hands or fingers at this rate. Many interesting performances such as dancing or martial arts involve rotational deformations that are much more rapid. We used three lighting directions, 15° from

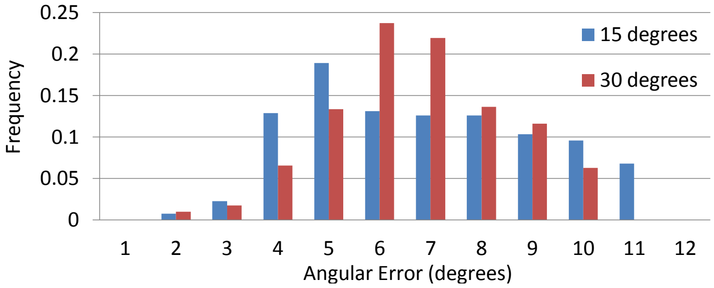

Fig. 3. Traditional photometric stereo using alternating white lights errs if the normal is changing. Here we show a histogram of the angular errors for estimated surface orientations using three alternating white lights, a 30fps camera, and a surface with 1Hz rotational deformation. We evenly sampled a range of surface orientations and rotational axes, for light directions 15° and 30° off the z axis. In both cases, we see a broad distribution of angular errors as high as 10°.

and evenly spaced around the z axis, and then repeated the simulations with the lighting 30° off the z axis. The surface normal for the middle frame was a vector $(0, 0, 1)$ pointing at the camera and rotated up or down (i.e. about the y axis) anywhere from -50° to 50°, in 10° increments. To simulate object motion, this normal rotated backward and forward 12° (for 30Hz rotation filmed with a 30fps camera) to generate the first and third measurements. We also changed the axis of rotation itself, using axes in the x-y plane, evenly spaced from 0° to 360° in 10° increments.

Figure 3 shows a histogram of the angular error between the true and computed surface normals for the middle frame. For the 15° off-axis lights, the mean and standard deviation of the angular error is 5.9° and 2.3°. For the 30° off-axis lights, the mean and standard deviation of the angular error is 5.7° and 1.7°. The computed normals are not simply averages of the observed ones; because of the varying lighting directions, even for a surface normal rotating in a plane, the computed orientation may not lie in the same plane. Orientation errors are also accompanied by reflectance errors. The mean albedo error (relative to the ground truth of 1.0) was 0.032 for light directions at 15° to the z axis, and 0.042 for 32°, with standard deviations of 0.039 and 0.050. Of course, these errors might change for different parameters. Regardless, they are a fundamental accuracy limit for dynamic photometric stereo with TMI if one ignores the time-varying normal.

RGB Component-Wise Imaging Models. Our second simulation investigated the errors from using a component-wise RGB camera for photometric stereo in the presence of crosstalk. In practice, such a system would alternate between red, green and blue light from a single direction in order to measure material reflectances, and spatially separated lights to measure surface orientation. We implemented the RGB component-wise imaging model using actual imaging data

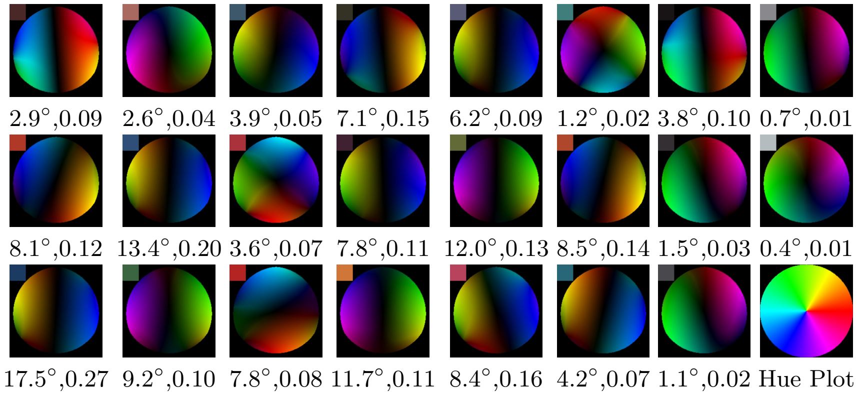

Fig. 4. Photometric stereo using colored lights with a simplified component-wise (R, G, B) model causes inaccurate normal and reflectance estimates. This simulation used red, green and blue lights 15° off the z-axis. These visualizations show hue plots of the surface normal directional error for spheres with colors corresponding to the Macbeth color checker patches. We show each patch's imaged color (inset squares), the maximum angular error (degrees) for estimated normals over the sphere, and the maximum albedo error (defined as the error for the computed normal's length, which should be 1.0). The white patch, used to fit the model, had negligible error.

for the Macbeth color checker and our LEDs and Flea2 camera, and simulated a stationary object (so these are ideal results). We took three pictures of the Macbeth chart illuminated by a single red, green, or blue LED, and computed the coefficient matrix $\mathbf{M}$ for each light, material and sensor combination. We let the color of the white Macbeth checker be $(1, 1, 1)$ and used the component-wise model to compute the color of each light, and then of all the checkers. We used the real-world imaging data to simulate Lambertian reflection off a sphere illuminated by a red, a green, and a blue light from 15° off the z axis, as before. To be conservative, we only simulated surface normals at angles less than 85° from all three lights.

Figure 4 shows that the angular orientation error using the component-wise model can be quite large. For the white patch, the simplified model works perfectly. The other patches show a fairly even spread of errors from nearly zero for the other grayscale patches to as high as 17°. The computed normals are all accurate at $(0, 0, 1)$ because the lights in these simulations are evenly spaced around the z axis, so the Lambertian shading terms are all equal at that one point. The clear axes in the error visualizations are due to our system having strong crosstalk between only two of the three color channels. These data show that using the component-wise imaging model leads to large surface orientation errors, especially at oblique angles. By contrast, our system, without sacrificing frame rate (i.e. still computing normals for every other input frame), computes accurate orientation despite significant crosstalk in two color channels.

5.2 Dynamic Photometric Stereo with Real Objects

Figure 5 shows three results using our system capture the shape and appearance of different moving and deforming objects. After computing surface normals, we reconstructed the surface geometry by integrating the surface normals, equivalent to solving a Poisson equation[12]. As the images show, we recovered the fine detail on the hands and glove (veins, wrinkles, etc.) well. The geometry for the creasing pillow is consistent despite sudden color changes. The mat is particularly interesting; the color texture is very complicated and prominent, yet barely detectable in the recovered geometry. The reader is encouraged to view the supplemental videos showing the motion in the input images, the reconstructed geometry, and the textured geometry for all three sequences. The hands rotate at roughly 1/6Hz, and the fingers curl even more rapidly.

6 Discussion

The fundamental goal of photometric stereo for moving, non-rigid objects is to estimate time-varying surface orientations. The prior art using TMI, however, assumes constant surface orientations across the frames used for the estimates, fundamentally limiting their accuracy. By contrast, our time and color multiplexed photometric stereo method is the first that is robust to changing surface orientations for non-rigid scenes. We have shown that for Lambertian surfaces and general imaging models, five images with appropriately chosen lighting are sufficient to recover the time-varying surface orientation for one frame. Our optimized implementation requires only three images. Because our method measures reflectance coefficients independently of the changing surface orientations, it preserves the key strength of colored lights for photometric stereo: an instantaneous orientation estimate in one frame, given known material reflectances.

Like the prior art, we use optical flow to align measurements from several video frames. Shading changes due to the deforming surfaces may complicate this alignment. Even in the ideal case of perfect optical flow, however, time-varying surface normals lead to errors for the prior art. By contrast, our method does not even require pixel accurate alignment. As long as surfaces of the same material are aligned to each other, the imaging coefficients are estimated correctly. Our implementation not only tolerates less than pixel-accurate alignment, but also makes the alignment more robust by fixing the apparent illumination for video in the camera red color channel. Using standard optical flow methods for that channel alone, we can directly align the frames used to measure material properties to the one with spatially separated lights for the orientation estimates. We need not assume linear motion across multiple frames, nor capture extra images to serve as optical flow key frames.

Our system is simple, consisting of an ordinary camera and LEDs with filters, yet captures detailed shapes of moving objects with complicated color textures. Like other three-source photometric stereo methods, it can err in the presence of non-Lambertian reflectance, interreflection, occlusions and mixed pixels. One might argue for a system with three completely isolated color channels (a red sensor that only responds to a red light, and so on). In addition to being difficult to implement in practice, such a design has another drawback: it cannot measure materials with no reflectance in one or more of the color channels. Our implementation suffers such limitations, but to a lesser extent. Saturated colors are not uncommon. Because our theory assumes a general imaging model with crosstalk, the five image solution could be used with broad-spectrum light colors and color sensors to capture the shapes of materials with a very wide range of colors (the constraint is that the three $\alpha_{\mathbf{k}}$ must be linearly independent).

References

1. Basri, R., Frolova, D.: A two-frame theory of motion, lighting and shape. In: IEEE Conference on Computer Vision and Pattern Recognition (2009)
2. Black, M., Anandan, P.: A framework for the robust estimation of optical flow. In: IEEE International Conference on Computer Vision. pp. 231–236 (1993)
3. Decker, B.D., Kautz, J., Mertens, T., Bekaert, P.: Capturing multiple illumination conditions using time and color multiplexing. In: IEEE Conference on Computer Vision and Pattern Recognition (2009)
4. Gretag Macbeth Color Management Solutions: http://www.gretagmacbeth.com
5. Hernández, C., Vogiatzis, G., Brostow, G.J., Stenger, B., Cipolla, R.: Non-rigid photometric stereo with colored lights. In: IEEE International Conference on Computer Vision. pp. 1–8 (2007)
6. Joshi, N., Kriegman, D.: Shape from varying illumination and viewpoint. In: IEEE International Conference on Computer Vision (2007)
7. Kontsevich, L.L., Petrov, A.P., Vergelskaya, I.S.: Reconstruction of shape from shading in color images. J. Opt. Soc. Am. A 11(3), 1047–1052 (March 1994)
8. Moses, Y., Shimshoni, I.: 3d shape recovery of smooth surfaces: Dropping the fixed viewpoint assumption. In: Asian Conference on Computer Vision (2006)
9. Nehab, D., Rusinkiewicz, S., Davis, J., Ramamoorthi, R.: Efficiently combining positions and normals for precise 3D geometry. ACM Trans. on Graphics 24(3), 536–543 (July 2005)
10. Petrov, A.: Light, color and shape. Cognitive Processes and their Simulation (in Russian) pp. 350–358 (1987)
11. Simakov, D., Frolova, D., Basri, R.: Dense shape reconstruction of a moving object under arbitrary, unknown lighting. In: IEEE International Conference on Computer Vision (2003)
12. Simchony, T., Chellappa, R., Shao, M.: Direct analytical methods for solving poisson equations in computer vision problems. IEEE Trans. on Pattern Analysis and Machine Intelligence 12(5), 435–446 (May 1990)
13. Vlasic, D., Peers, P., Baran, I., Debevec, P., Popović, J., Rusinkiewicz, S., Matusik, W.: Dynamic shape capture using multi-view photometric stereo. ACM Trans. on Graphics 28(5) (2009)
14. Weise, T., Leibe, B., Gool, L.V.: Fast 3d scanning with automatic motion compensation. In: IEEE Conference on Computer Vision and Pattern Recognition (2007)
15. Wenger, A., Gardner, A., Tchou, C., Unger, J., Hawkins, T., Debevec, P.: Performance relighting and reflectance transformation with time-multiplexed illumination. ACM Trans. on Graphics 24(3), 756–764 (2005)
16. Woodham, R.: Photometric method for determining surface orientation from multiple images. Optical Engineering 19(1), 139–144 (1980)

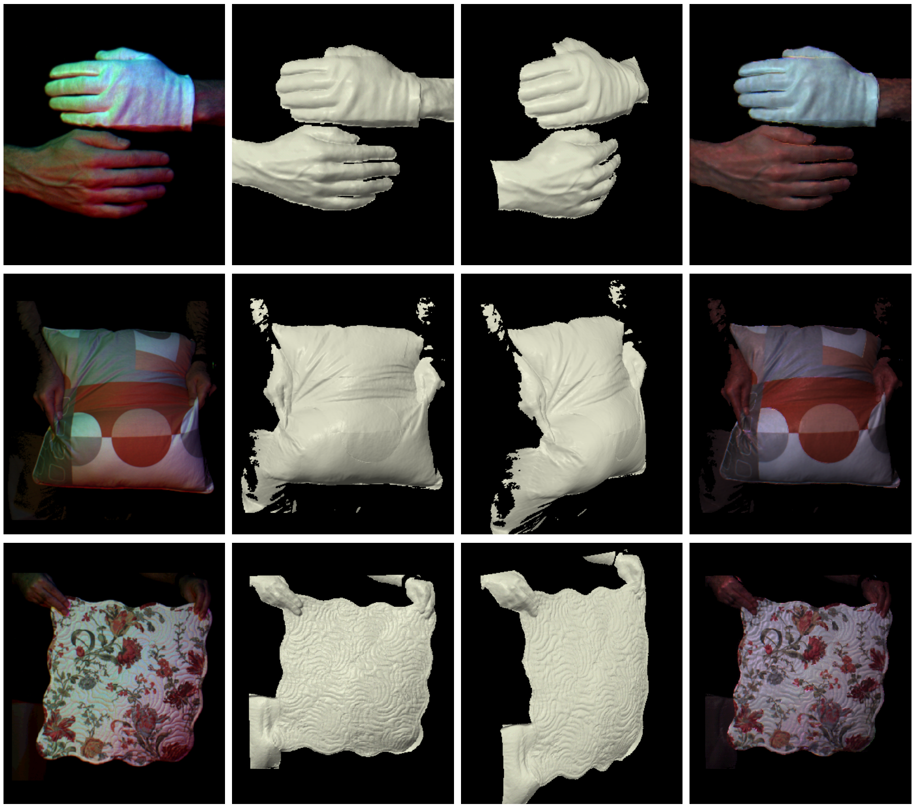

Fig. 5. Results from our photometric stereo system. From top to bottom, the rows are (1) a video frame illuminated with spatially separated red, green and blue LEDs (2) reconstructed geometry (3) geometry rendered from a new view, and (4) geometry rendered from the camera view and textured with measured appearance data. The hands are rotating around the axis of the arms, the pillow is being creased, and the mat is being waved. The pillow shows that we are computing consistent normals despite changing material colors. The artifacts at color edges are due to resampling during image alignment. We recover the fine quilted surface detail of the mat well despite its colorful pattern. The supplemental material video shows the entire input and output video sequences.