

Photometric Stereo with Small Angular Variations

Jian Wang[†], Yasuyuki Matsushita[‡], Boxin Shi[§], and Aswin C. Sankaranarayanan[†]

[†]ECE Department, Carnegie Mellon University, Pittsburgh, PA, USA

[‡]Osaka University, Osaka, Japan

[§]Singapore University of Technology and Design, Singapore

Abstract

Most existing successful photometric stereo setups require large angular variations in illumination directions, which results in acquisition rigs that have large spatial extent. For many applications, especially involving mobile devices, it is important that the device be spatially compact. This naturally implies smaller angular variations in the illumination directions. This paper studies the effect of small angular variations in illumination directions to photometric stereo. We explore both theoretical justification and practical issues in the design of a compact and portable photometric stereo device on which a camera is surrounded by a ring of point light sources. We first derive the relationship between the estimation error of surface normal and the baseline of the point light sources. Armed with this theoretical insight, we develop a small baseline photometric stereo prototype to experimentally examine the theory and its practicality.

1. Introduction

Size, weight, and power (SWaP) are key factors in designing a practical 3D acquisition system. The popularity of commercial depth cameras such as the Kinect [14], which is based on time-of-flight technology, and RealSense [10], which is based on stereo and active illumination, can be attributed to their careful SWaP consideration. In contrast to this, 3D acquisition based on photometric stereo is yet to gain widespread commercial adoption. There are many advantages to use photometric stereo — all arising from its ability to compute surface orientation at the same resolution as the input images [20, 18], which is achieved by determining surface normal per pixel from shading variations observed under varying lightings. Reliable surface normal estimates are obtained when the light sources have a large angular spread [4]; however, lighting with a large angular spread necessarily requires a large space especially when imaging macroscopic objects, and hence, does not satisfy small SWaP.

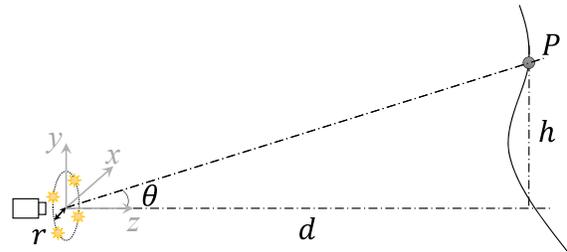


Figure 1. A compact photometric stereo system with a light ring of radius r and a camera placed at the center of the ring. We seek to image a scene whose depth d is significantly larger than the radius of the light ring, i.e., $d \gg r$.

This paper deals with photometric stereo under small angular variations. Our envisioned acquisition setup is as follows. A camera is surrounded by a small light ring. It is used to capture the surface normal of a scene whose spatial extent is significantly larger than the system itself. We define the baseline as the greatest distance between every two lights, which is the diameter of the light ring here. To fulfill the small SWaP requirement, we want the baseline to be as small as possible.

1.1. Problem setup

In this paper, we analyze the estimation error with regard to baseline and other system parameters like the number of lights, camera noise level and scene location. Our problem formulation is as follows.

As illustrated in Fig. 1, our setup consists of n identical point light sources distributed uniformly on a circle of radius r , centered at the camera. We obtain n images, one for illumination under each light source. The camera is assumed to make intensity measurements corrupted by additive noise with mean 0, variance σ^2 , and *i.i.d* both across pixels and across images. Finally, the baseline of the system is small, which translates to the condition $r \ll d$.

We consider a Lambertian scene point P at the location $\mathbf{p} = [0, h, d]^T \in \mathbb{R}^3$ with a surface normal $\mathbf{n} \in \mathbb{R}^3$, $\|\mathbf{n}\|_2 = 1$ and diffuse albedo ρ . When P is illuminated by a point light source at the location $\mathbf{s} \in \mathbb{R}^3$, its

intensity is given as

$$\hat{i} = i + \Delta i_g = \frac{\mathbf{l}^\top (\rho \mathbf{n})}{\|\mathbf{l}\|^3} + \Delta i_g, \quad (1)$$

where $\mathbf{l} = \mathbf{s} - \mathbf{p}$ is the illumination direction, $\|\mathbf{l}\|^3$ comes from normalization and light fall-off [2], light source intensity at a unit distance is assumed as 1, and Δi_g is the measurement noise. In the absence of measurement noise, the baseline of the system can be infinitely small as long as it is non-zero.¹ However, in the presence of measurement noise, the variations in shading become less obvious with decreasing baseline, and thus the recovered surface normals will have larger errors. We note here that the variance of Δi_g , σ^2 , accounts for the camera response function. Given three or more intensity observations obtained under varying lightings $\mathbf{l}_1, \dots, \mathbf{l}_n$ located at $\mathbf{s}_1, \dots, \mathbf{s}_n$, respectively, $\hat{\mathbf{b}}$ which is the estimate of the albedo-scaled surface normal $\rho \mathbf{n}$ is given as a least-squares approximate solution as

$$\hat{\mathbf{b}} = (\mathbf{L}\mathbf{L}^\top)^{-1}\hat{\mathbf{L}}\mathbf{i}, \quad (2)$$

where

$$\mathbf{L} = \begin{bmatrix} \frac{\mathbf{l}_1}{\|\mathbf{l}_1\|^3} & \frac{\mathbf{l}_2}{\|\mathbf{l}_2\|^3} & \dots & \frac{\mathbf{l}_n}{\|\mathbf{l}_n\|^3} \end{bmatrix},$$

and

$$\hat{\mathbf{i}} = [\hat{i}_1 \quad \hat{i}_2 \quad \dots \quad \hat{i}_n]^\top.$$

The matrix \mathbf{L} is referred to as the light matrix or the light calibration matrix. This matrix is often estimated via a calibration procedure, *i.e.*, placing known targets at suitable depths in the scene.

Traditional photometric stereo systems assume distant lighting — which implies that the light matrix \mathbf{L} is identical at all pixels/scene points. In contrast, it is extremely important in the small angular case to account for variations of the light matrix as the scene point P of interest varies.

1.2. Main results

Given $\mathbf{b} = \rho \mathbf{n}$ and the estimated $\hat{\mathbf{b}}$ in (2), the contributions of this paper are in analyzing the expected error $E[\|\hat{\mathbf{b}} - \mathbf{b}\|^2]$. We summarize these results below.

Theorem 1 *In the small-baseline scenario, *i.e.*, $d \gg r$,*

$$E_{\Delta i_g}[\|\hat{\mathbf{b}} - \rho \mathbf{n}\|^2] = \sigma^2 (d^2 + h^2)^3 \frac{2(2d^2 + h^2)}{nr^2 d^2}, \quad (3)$$

or equivalently,

$$E_{\Delta i_g}[\|\hat{\mathbf{b}} - \rho \mathbf{n}\|^2] = \sigma^2 (d^2 + h^2)^2 \frac{4\pi}{n\Omega \frac{2 \cos \theta}{1 + \cos^2 \theta}}, \quad (4)$$

¹When the baseline is zero, the problem reduces to shape-from-shading which is known to be severely ill-posed [9].

where Ω is the solid angle subtended at scene point by the light ring and $\theta = \text{atan}(h/d)$ is the angle between surface normal of the light ring and the line from center of light ring to the scene point.

Theorem 1 predicts the accuracy of the estimate in (2) as a function of measurement noise and the location of the scene point. Intuitively, we expect that the estimation error increases as we move away from the optical axis of the camera since the solid angle subtended by the light ring reduces to a 1D line with increasing height h . This is reflected implicitly in (3) as well as in (4) where the dependence on the subtended solid angle is explicitly characterized. The proof of Theorem 1 is presented in Section 3.

Also note that the estimate $\hat{\mathbf{b}}$ is dependent on the light matrix \mathbf{L} which relies on knowledge of the scene point P and hence, depth d . In practice, this matrix is estimated via a calibration procedure. So, this reduces to approximate the unknown depth d with a value \hat{d} . Our second contribution is bounding the error due to use of incorrect depth in the estimation of $\hat{\mathbf{b}}$.

Theorem 2 (Sensitivity to calibration error) *Consider a scene point that is actually at depth d , but we assume that it is at depth \hat{d} ($d, \hat{d} \gg r$). Further, if we assume the surface normal at this point, \mathbf{n} , to be uniformly distributed, then*

$$E_{\mathbf{n}}[\|\hat{\mathbf{b}} - \rho \mathbf{n}\|^2] = \frac{\rho^2}{3} (\lambda - 1)^2 (2(\lambda^2 + \lambda + 1)^2 + (\lambda + 1)^2), \quad (5)$$

where $\lambda = \hat{d}/d$.

The proof of Theorem 2 is presented in Section 4.

From Theorem 1, we observe that a practical photometric stereo system faces a trade-off between small SWaP and the accuracy. From Theorem 2, we observe that the decrease in estimation accuracy caused by depth mismatch is gradual on each side of the calibrated depth, and there is a depth range where error is tolerable, which relaxes the exact placement of the target scene. Based on these observations, we develop a small baseline photometric stereo prototype to experimentally examine the theory and its practicality.

2. Related work

Many practical systems based on photometric stereo have been proposed since the early work of Woodham [20] and Silver [18]. For example, Hernández *et al.* [7] used colored lights to capture images under multiple illumination directions simultaneously; this enabled estimation of surface normals of moving objects under a snapshot acquisition system. Vlasic *et al.* [19] built a large light stage to capture surface normal of actors from nearly 360°. However, both systems are too large to be portable. Compact acquisition devices were proposed by Higo *et al.* [8], Zhou

and Tan [22], Zhou *et al.* [23], Shiradkar *et al.* [17], and Johnson *et al.* [11]; however, all of these devices were used to image scenes with spatial extent that was comparable to the size of the light path [8] or light ring [22, 23, 17, 11]. In contrast, the primary focus of this paper is to understand a more extreme setup where we seek to use a portable device to capture scenes that are much larger than the device (see Fig. 1). Jones *et al.* [12] used a device similar to ours to capture expressions of a human face, but did not provide a detailed error analysis including a characterization of the dependence of the normal estimation error on the baseline.

Closely related to this paper is the work of Drbohlav and Chantler [4] where an error analysis of Lambertian photometric stereo is presented. However, a key difference is the assumption of distant lighting which is only applicable to large baseline systems. The analysis in [4] can be seen as a special case when $h = 0$ in our setting.

Our small baseline photometric stereo setup is similar to multi-flash cameras [16, 5]. However, the goal of such cameras is to recover depth discontinuities. Further, our focus is largely a theoretical analysis of the estimation error which was not considered in these papers.

3. Error due to measurement noise

We analyze the error in estimation of albedo-scaled surface normals with regard to the baseline and the solid angle subtended at scene point by the light ring, and in the process, provide the proof for Theorem 1.

There are many sources of noise in the process of imaging including photon, thermal, and dark-current noise. For simplicity, we model the measurement noise as zero mean and bounded variance. Specifically, we assume an intensity measurement made at any pixel can be written as

$$\hat{i} = i + \Delta i_g, \quad (6)$$

where i is the noise-free measurement and Δi_g is the additive measurement noise with mean 0 and variance σ^2 . We assume the measurement noise Δi_g is *i.i.d.* across different pixels in an image and across images.

Given the albedo-scaled surface normal $\mathbf{b} = \rho \mathbf{n}$ and its estimate $\hat{\mathbf{b}}$ in (2), we define the error metric $e_{l_2} = \|\hat{\mathbf{b}} - \mathbf{b}\|^2$. The expected value of e_{l_2} can be written as

$$E[e_{l_2}] = \sigma^2 \text{trace} [(\mathbf{L}\mathbf{L}^\top)^{-1}], \quad (7)$$

following the derivation of (33) in [4].

3.1. Dependence of error on baseline

We derive the error expectation with regard to baseline and other system parameters like measurement noise level, number of lights and the 3D location of the target.

From (7), we observe that $E[e_{l_2}]$ is the product of the variance of measurement noise, σ^2 , and trace $[(\mathbf{L}\mathbf{L}^\top)^{-1}]$.

Here, we further express trace $[(\mathbf{L}\mathbf{L}^\top)^{-1}]$ as a function of the radius of light ring r , the number of lights n , and scene point's location by using the following four steps.

Step 1 — Location of light sources. Because lights are uniformly distributed on a ring, their locations can be written as

$$\mathbf{S} = [\mathbf{s}_1 \quad \mathbf{s}_2 \quad \dots \quad \mathbf{s}_n] = \begin{bmatrix} r \cos(\frac{2\pi}{n}1) & \dots & r \cos(\frac{2\pi}{n}n) \\ r \sin(\frac{2\pi}{n}1) & \dots & r \sin(\frac{2\pi}{n}n) \\ 0 & \dots & 0 \end{bmatrix}, \quad (8)$$

where each column \mathbf{s}_i is 3D location of each light.

Step 2 — Expression for \mathbf{L} . Given the definition of \mathbf{S} , we can derive the light matrix \mathbf{L} as

$$\begin{aligned} \mathbf{L} &= [\mathbf{l}_1 \quad \mathbf{l}_2 \quad \dots \quad \mathbf{l}_n] \\ &= \begin{bmatrix} \frac{\mathbf{s}_1 - \mathbf{p}}{\|\mathbf{s}_1 - \mathbf{p}\|^3} & \frac{\mathbf{s}_2 - \mathbf{p}}{\|\mathbf{s}_2 - \mathbf{p}\|^3} & \dots & \frac{\mathbf{s}_n - \mathbf{p}}{\|\mathbf{s}_n - \mathbf{p}\|^3} \end{bmatrix} \\ &\stackrel{(a)}{\approx} \frac{1}{(d^2 + h^2)^{\frac{3}{2}}} [\mathbf{s}_1 - \mathbf{p} \quad \mathbf{s}_2 - \mathbf{p} \quad \dots \quad \mathbf{s}_n - \mathbf{p}] \\ &= \frac{1}{(d^2 + h^2)^{\frac{3}{2}}} \left[\mathbf{S} - \begin{bmatrix} 0 \\ h \\ d \end{bmatrix} \mathbf{1}^\top \right], \end{aligned} \quad (9)$$

where $\stackrel{(a)}{\approx}$ uses the small baseline criteria, $r \ll d$, so that $\|\mathbf{s}_i - \mathbf{p}\| \approx (d^2 + h^2)^{\frac{1}{2}}$.

Step 3 — Expression for $\mathbf{L}\mathbf{L}^\top$ and its inverse.

$$\begin{aligned} \mathbf{L}\mathbf{L}^\top &= \frac{1}{(d^2 + h^2)^3} \left[\mathbf{S} - \begin{bmatrix} 0 \\ h \\ d \end{bmatrix} \mathbf{1}^\top \right] \left[\mathbf{S} - \begin{bmatrix} 0 \\ h \\ d \end{bmatrix} \mathbf{1}^\top \right]^\top \\ &= \frac{1}{(d^2 + h^2)^3} \begin{bmatrix} \frac{n}{2}r^2 & 0 & 0 \\ 0 & \frac{n}{2}r^2 + nh^2 & ndh \\ 0 & ndh & nd^2 \end{bmatrix} \end{aligned} \quad (10)$$

$$(\mathbf{L}\mathbf{L}^\top)^{-1} = (d^2 + h^2)^3 \begin{bmatrix} \frac{2}{nr^2} & 0 & 0 \\ 0 & \frac{2}{nr^2} & -\frac{2h}{ndr^2} \\ 0 & -\frac{2h}{ndr^2} & \frac{r^2 + 2h^2}{nr^2d^2} \end{bmatrix} \quad (11)$$

Step 4 — Expression for $\text{trace}(\mathbf{L}\mathbf{L}^\top)^{-1}$.

$$\begin{aligned}
& \text{trace} [(\mathbf{L}\mathbf{L}^\top)^{-1}] \\
&= (d^2 + h^2)^3 \left(\frac{2}{nr^2} + \frac{2}{nr^2} + \frac{r^2 + 2h^2}{nr^2 d^2} \right) \\
&= (d^2 + h^2)^3 \frac{r^2 + 4d^2 + 2h^2}{nr^2 d^2} \\
&\stackrel{(a)}{\approx} (d^2 + h^2)^3 \frac{2(2d^2 + h^2)}{nr^2 d^2},
\end{aligned} \tag{12}$$

where $\stackrel{(a)}{\approx}$ follows that $r \ll d$.

Replacing (12) in (7), we obtain

$$\boxed{E[e_{l_2}] = \sigma^2 (d^2 + h^2)^3 \frac{2(2d^2 + h^2)}{nr^2 d^2}}. \tag{13}$$

This proves the first part of Theorem 1.

3.2. Dependence of error on solid angle

We now investigate the dependence of the estimate on the solid angle subtended by the light ring at a scene point. Intuitively, the solid angle is more meaningful from a geometric perspective since it better captures the light configuration as seen at a scene point (see Fig. 2).

The solid angle Ω subtended by the light ring at a scene point $\mathbf{p} = (0, h, d)^\top$ can be written as

$$\Omega = \int_0^{2\pi} \int_0^r \frac{ld}{(l^2 + h^2 + d^2 + 2hl \sin \phi)^{\frac{3}{2}}} dl d\phi, \tag{14}$$

but the closed form of the double integration is analytically intractable [6]. We instead approximate it as

$$\Omega \approx \frac{\pi r^2 \cos \theta}{h^2 + d^2}, \tag{15}$$

where we employ two approximations: first, the numerator, which is the area of an ellipse with semi-major axis r and semi-minor axis $r \cos \theta$, is approximated as $\pi r^2 \cos \theta$, and second, the denominator, which is the distance from the point to the ellipse, as the distance of the point \mathbf{p} to the center of the light ring.

Substituting (15) in (13), we obtain:

$$\boxed{E[e_{l_2}] = \sigma^2 (d^2 + h^2)^2 \frac{4\pi}{n\Omega} \frac{2 \cos \theta}{1 + \cos^2 \theta}}. \tag{16}$$

This provides the proof for the second part of Theorem 1. In the expression, factor $(d^2 + h^2)^2$ is caused by light fall-off; $\frac{2 \cos \theta}{1 + \cos^2 \theta}$ decreases from 1 to 0 when θ increases from 0 to $\pi/2$, thus it can be viewed as discount rate of the solid angle Ω ; in other words, error is inversely proportional to “discounted” solid angle ($\Omega \frac{2 \cos \theta}{1 + \cos^2 \theta}$). In particular, it is noteworthy that the error is not just inversely proportional to Ω , but also dependent on the angle θ .

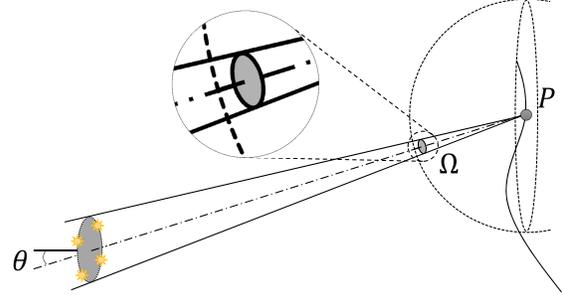


Figure 2. Solid angle subtended at point P by the light ring.

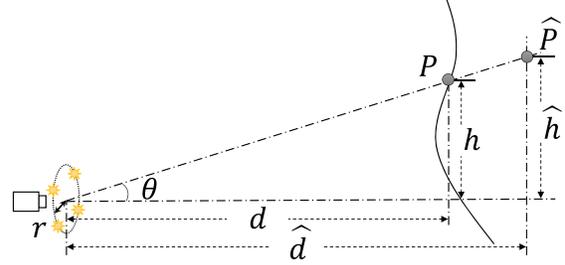


Figure 3. Illustration of error due to mis-calibration. Surface normal of point P is computed using light matrix of point \hat{P} .

4. Error due to incorrect calibration

In this section, we derive the error due to our lack of knowledge of the exact depth at which the target is placed. Recall that, our results hold for scenarios where the distant lighting assumption is violated and hence, we need to account for spatial variations in the direction of the incident illumination. However, this requires knowledge of the scene depth which is often the goal of photometric stereo. While we can often simply use an approximate value for depth and hope to get reasonable results, we provide a theoretical characterization of the error incurred due to this “mis-calibration”.

Suppose that the true scene point is at a location P and we assume it to be, incorrectly, at \hat{P} (see Fig. 3). When scene point is at P , intensities \mathbf{i} are $\mathbf{L}^\top \mathbf{b}$ where \mathbf{L} is the light matrix. When using light matrix associated with the point \hat{P} to compute the albedo-scaled surface normal, according to (2), we obtain the estimate

$$\hat{\mathbf{b}} = (\hat{\mathbf{L}}\hat{\mathbf{L}}^\top)^{-1} \hat{\mathbf{L}}^\top \mathbf{b}, \tag{17}$$

where $\hat{\mathbf{L}}$ is light matrix of point \hat{P} .

4.1. The hypothesized light ring

We first introduce the so-called “hypothesized light ring” and use it to obtain an approximation for the term $(\hat{\mathbf{L}}\hat{\mathbf{L}}^\top)^{-1} \hat{\mathbf{L}}^\top$. Subsequently, we derive expected value of e_{l_2} both in the absence and presence of measurement noise.

Specifically, we assume that the light ring as seen at a point P can be replaced by a “hypothesized light ring”,

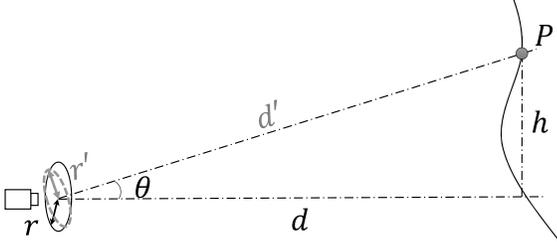


Figure 4. Illustration of the hypothesized light ring, which is orthogonal to the line from the camera to the scene point.

with radius r' and distance d' , that is orthogonal to the line from camera to the scene point (see Fig. 4). Clearly, $d' = d/\cos\theta$. An expression for r' is obtained by ensuring that the original light ring and the hypothesized light ring should generate equal error. Specifically, the expression in (13) for a scene point $(0, h, d)^\top$ with a light ring of radius r should be identical to that of a scene point at $(0, 0, d')^\top$ and a light ring with radius r' . This gives us

$$r' = \left(\frac{2\cos^2\theta}{1+\cos^2\theta} \right)^{\frac{1}{2}} r. \quad (18)$$

In contrast to derivation by foreshortening ($r\cos\theta$), the derived error estimation by r' matches much better with simulations.

The light matrix of hypothesized lights \mathbf{L} can be expressed as

$$\mathbf{L} = \frac{1}{q^3} \mathbf{R} \left[\mathbf{S}(r') - \begin{bmatrix} 0 \\ 0 \\ q \end{bmatrix} \mathbf{1}^\top \right] \quad (19)$$

where q is $(d^2 + h^2)^{\frac{1}{2}}$, $\mathbf{S}(r')$ is the hypothesized lights' locations which is obtained by changing r of (8) to r' , and \mathbf{R} is the coordinate rotation matrix given as

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{bmatrix}.$$

We can now express $\widehat{\mathbf{L}}\widehat{\mathbf{L}}^\top$ as

$$\widehat{\mathbf{L}}\widehat{\mathbf{L}}^\top = \frac{1}{\widehat{q}^6} \mathbf{R} \begin{bmatrix} \frac{n}{2}r'^2 & 0 & 0 \\ 0 & \frac{n}{2}r'^2 & 0 \\ 0 & 0 & n\widehat{q}^2 \end{bmatrix} \mathbf{R}^\top \quad (20)$$

where \widehat{q} is $(\widehat{d}^2 + \widehat{h}^2)^{\frac{1}{2}}$, and $\widehat{\mathbf{L}}\widehat{\mathbf{L}}^\top$ as

$$\widehat{\mathbf{L}}\widehat{\mathbf{L}}^\top = \frac{1}{\widehat{q}^3\widehat{q}^3} \mathbf{R} \begin{bmatrix} \frac{n}{2}r'^2 & 0 & 0 \\ 0 & \frac{n}{2}r'^2 & 0 \\ 0 & 0 & n\widehat{q}\widehat{q} \end{bmatrix} \mathbf{R}^\top. \quad (21)$$

Combining (20) and (21) together, we have $(\widehat{\mathbf{L}}\widehat{\mathbf{L}}^\top)^{-1}\widehat{\mathbf{L}}\mathbf{L}^\top$ equal to

$$\frac{\widehat{d}^3}{d^3} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos^2\theta + \sin^2\theta\frac{d}{\widehat{d}} & -\sin\theta\cos\theta(1-\frac{d}{\widehat{d}}) \\ 0 & -\sin\theta\cos\theta(1-\frac{d}{\widehat{d}}) & \sin^2\theta + \cos^2\theta\frac{d}{\widehat{d}} \end{bmatrix}.$$

Denoting $\mathbf{D} = (\widehat{\mathbf{L}}\widehat{\mathbf{L}}^\top)^{-1}\widehat{\mathbf{L}}\mathbf{L}^\top$, (17) can be written as $\widehat{\mathbf{b}} = \mathbf{D}\mathbf{b}$, and $e_{l_2} = \|\widehat{\mathbf{b}} - \mathbf{b}\|^2 = (\widehat{\mathbf{b}} - \mathbf{b})^\top (\widehat{\mathbf{b}} - \mathbf{b}) = (\mathbf{D}\mathbf{b} - \mathbf{b})^\top (\mathbf{D}\mathbf{b} - \mathbf{b}) = \mathbf{b}^\top (\mathbf{D} - \mathbf{I})^\top (\mathbf{D} - \mathbf{I})\mathbf{b}$.

To obtain an expression for e_{l_2} that is independent of the surface normal, we assume a uniform distribution on the surface normals, described in its Euler angles. Specifically, letting

$$\mathbf{n} = [\rho\sin\Theta\cos\Phi, \rho\sin\Theta\sin\Phi, \rho\cos\Theta]^\top,$$

with probability density function $f_{\Theta,\Phi} = \frac{1}{4\pi}\sin(\Theta)$, Θ from 0 to π and Φ from 0 to 2π , and $\lambda = \frac{\widehat{d}}{d}$, we can compute

$$E_{\Phi,\Theta}[e_{l_2}] = \frac{1}{3}\rho^2(\lambda-1)^2(2(\lambda^2+\lambda+1)^2+(\lambda+1)^2). \quad (22)$$

This completes the proof for Theorem 2. The result is quite surprising because it only relates to λ and does not depend on r , n and h .

Finally, in the presence of measurement noise, where the image intensities $\widehat{\mathbf{i}}$ is given as $(\mathbf{L}^\top\mathbf{b} + \Delta\mathbf{i}_g)$, we can write

$$\widehat{\mathbf{b}} = (\widehat{\mathbf{L}}\widehat{\mathbf{L}}^\top)^{-1}\widehat{\mathbf{L}}(\mathbf{L}^\top\mathbf{b} + \Delta\mathbf{i}_g) = \mathbf{D}\mathbf{b} + (\widehat{\mathbf{L}}\widehat{\mathbf{L}}^\top)^{-1}\widehat{\mathbf{L}}\Delta\mathbf{i}_g. \quad (23)$$

Now, the expression for $E[e_{l_2}]$ can be derived as

$$\begin{aligned} E[e_{l_2}] &= E_{\Theta,\Phi,\Delta\mathbf{i}_g} \left[(\widehat{\mathbf{b}} - \mathbf{b})^\top (\widehat{\mathbf{b}} - \mathbf{b}) \right] \\ &= E_{\Theta,\Phi,\Delta\mathbf{i}_g} \left[\left((\mathbf{D} - \mathbf{I})\mathbf{b} + (\widehat{\mathbf{L}}\widehat{\mathbf{L}}^\top)^{-1}\widehat{\mathbf{L}}\Delta\mathbf{i}_g \right)^\top \right. \\ &\quad \left. \left((\mathbf{D} - \mathbf{I})\mathbf{b} + (\widehat{\mathbf{L}}\widehat{\mathbf{L}}^\top)^{-1}\widehat{\mathbf{L}}\Delta\mathbf{i}_g \right) \right] \\ &= E_{\Theta,\Phi} \left[\mathbf{b}^\top (\mathbf{D} - \mathbf{I})^\top (\mathbf{D} - \mathbf{I})\mathbf{b} \right] + \\ &\quad 2E_{\Theta,\Phi,\Delta\mathbf{i}_g} \left[\mathbf{b}^\top (\mathbf{D} - \mathbf{I})^\top (\widehat{\mathbf{L}}\widehat{\mathbf{L}}^\top)^{-1}\widehat{\mathbf{L}}\Delta\mathbf{i}_g \right] + \\ &\quad E_{\Delta\mathbf{i}_g} \left[\left((\widehat{\mathbf{L}}\widehat{\mathbf{L}}^\top)^{-1}\widehat{\mathbf{L}}\Delta\mathbf{i}_g \right)^\top (\widehat{\mathbf{L}}\widehat{\mathbf{L}}^\top)^{-1}\widehat{\mathbf{L}}\Delta\mathbf{i}_g \right] \\ &\stackrel{(a)}{=} E_{\Theta,\Phi} \left[\mathbf{b}^\top (\mathbf{D} - \mathbf{I})^\top (\mathbf{D} - \mathbf{I})\mathbf{b} \right] + \sigma^2 \text{trace} \left[(\widehat{\mathbf{L}}\widehat{\mathbf{L}}^\top)^{-1} \right] \end{aligned} \quad (24)$$

where $\stackrel{(a)}{=}$ follows from the second term being 0 and (7). The final error is a linear combination of error due to incorrect placement (the first term in the sum), denoted as E_1 below, and error due to measurement noise (the second term). E_1 is only related to the albedo ρ and the ratio λ , and not related to n , r and h . When $\widehat{d} = d$, i.e., $\lambda = 1$, E_1 is zero. When

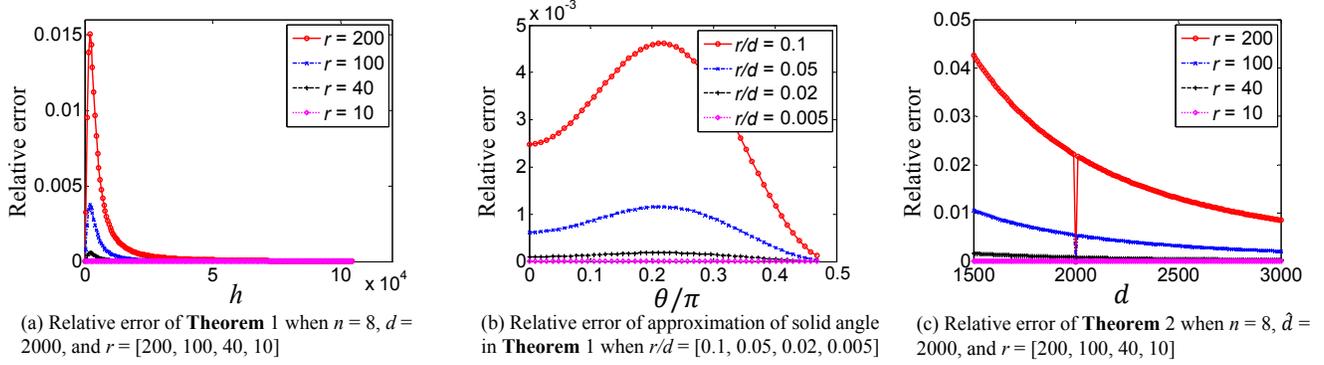


Figure 5. Relative error of the approximations in theorems.

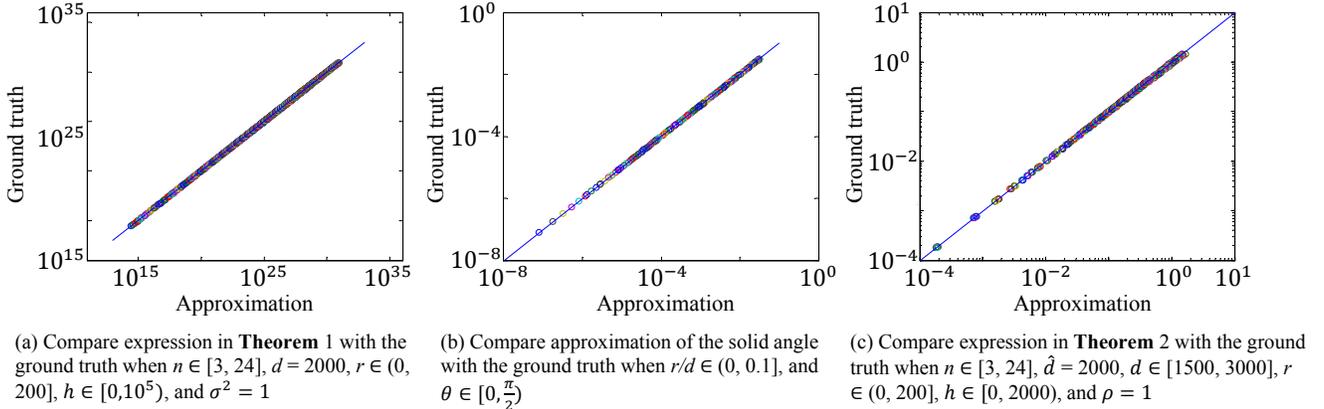


Figure 6. Comparison of the approximations in theorems with the ground truth. Solid line is “ $y = x$.”

\hat{d} deviates from d , E_1 increases gradually. Such analysis is particularly useful to compute a range of depth values where error is tolerable.

Limitations of Theorems 1 and 2. It is often more meaningful to analyze the accuracy of a photometric stereo system in terms of angular error $e_\angle = \arccos \frac{\hat{\mathbf{b}}^\top \mathbf{b}}{\|\hat{\mathbf{b}}\| \|\mathbf{b}\|}$ for surface normal estimates. However, while e_\angle is physically meaningful, analyzing it analytically is significantly harder than e_{l_2} which enjoys closed-form expressions. A second limitation is that we employ approximations in the derivation of both theorems. Next, we show using a wide range of simulations that in spite of these approximations, the error expressions are extremely precise.

Verification of the approximations. We compare the theoretical predictions of expected error in Theorems 1 and 2 to simulation results. Recall that our system is completely defined using five parameters: the radius of the light ring, r ; the number of point light sources, n ; the depth of the scene point, d ; the height of the scene point h ; and the variance of measurement noise, σ^2 . In addition to these, \hat{d} is the depth

at which we calibrate the light sources. We note that unless otherwise stated it is to be taken that $\hat{d} = d$.

Given a system configuration, denoted by the values of $\{r, n, d, h, \sigma^2\}$, we compute the expected error from the expressions in **Theorem 1** as well as from simulations. Using the simulations as ground truth, we compute the relative error defined as

$$\frac{|\text{ground truth} - \text{approximation}|}{\text{ground truth}}$$

Figure 5(a) verifies **Theorem 1** by comparing relative error as a function of h and r for fixed values of $d = 2000$, and $n = 8$ (σ^2 does not matter here). Figure 5(b) verifies the approximation of solid angle by (15) as a function of r/d and θ . Figure 5(c) verifies **Theorem 2**, assuming no measurement noise, by presenting accuracy in estimating calibration error as a function of the true depth d from 1500 to 3000 when we have incorrectly calibrated lights at $\hat{d} = 2000$. As expected, when $d = \hat{d} = 2000$, the error due to mis-calibration is zero.

Next, in Fig. 6, we show scatter plots that compare the ground truth to the theoretical expressions in the theorems. The proximity of the scatter plot to the “ $y = x$ ” line in-

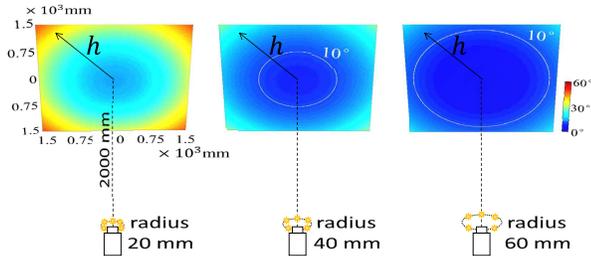


Figure 7. Mean angular error maps when $\sigma^2 = 2$, $n = 8$, $r = \{20mm, 40mm, 60mm\}$, and the true depth and calibrated depth both equal to $2000mm$. Notice how the angular error is radially symmetric — a consequence of the circular light ring. Angular error of the region enclosed by white line is smaller than 10° .

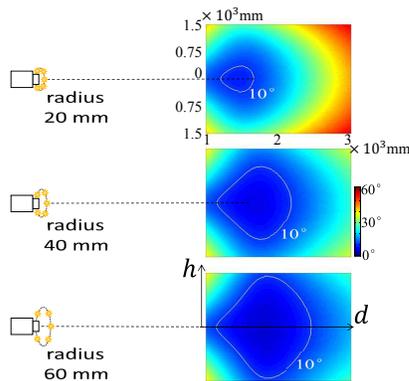


Figure 8. Mean angular error maps when $\sigma^2 = 2$, $n = 8$, $r = \{20mm, 40mm, 60mm\}$. The system is calibrated for a depth $\hat{d} = 2000mm$ while we vary the true depth $d \in [1000mm, 3000mm]$. The plots shown are a function of d (horizontal axis) and h (vertical axis). Angular error of the region enclosed by white line is smaller than 10° .

indicates that our theoretical predictions are very precise and broadly independent of system parameters.

5. Experiments

We showcase the predictions of the Theorems 1 and 2 by building an experimental prototype and comparing its performance to our theoretical predictions.

Choice of baseline. Before we build our prototype with a fixed baseline, it is first instructive to look at the theoretical predictions. Let us consider a system with eight lights, $n = 8$, and measurement noise with variance $\sigma^2 = 2$. We assume calibration for a depth $\hat{d} = 2000mm$. We vary the radius of the light ring $r \in \{20mm, 40mm, 60mm\}$ and consider two scenarios. First, in Fig. 7, we assume that the scene is at the calibrated depth $d = \hat{d} = 2000mm$ and look at mean angular error as a function of h . Second, in Fig. 8, we vary the true depth $d \in [1000mm, 3000mm]$ when the system is calibrated for $\hat{d} = 2000mm$ and plot the error

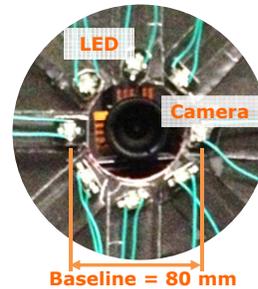


Figure 9. Our portable surface normal sensor with a light ring of radius $r = 40mm$ (diameter of $80mm$).

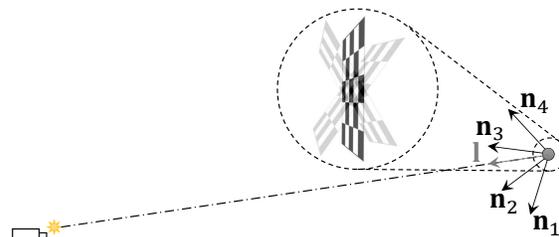


Figure 10. Light calibration method.

as a function of both d and h . In both cases, we mark the 10° equi-error contour in white. An important observation is that around $r = 40mm$ we get a sufficiently large region where the error is smaller than 10° . This suggests that we can have a system that can be fit onto a smart-phone.

Prototype. We used a Point Grey camera FL3-U3-13E4C-C and Cree XLamp XM-L LEDs to build a small baseline photometric stereo device (see Fig. 9) with the goal of recovering surface normals and depth maps of scenes placed nearly $2000mm$ away. We fixed the number of light sources on the ring as 8 for all experiments. Note that the number of lights in the rings controls the trade-off between accuracy ($1/n$ as detailed in Theorem 1) and acquisition time (proportional to n). While more lights lead to less error, they also result in longer acquisition time.

Light calibration. Due to the small baseline, precise calibration of the light matrix is very important. We observed that the traditional method of using specular and diffuse spheres provided calibration results were not sufficiently accurate. To resolve this, we propose a novel light calibration technique using a diffuse checkerboard pattern. Figure 10 illustrates our light calibration method. We use the Camera Calibration Toolbox [1] for estimating the location and surface normal of the checkerboard pattern. By observing the checkerboard at a certain 3D location under a certain lighting condition, we can recover light matrix by varying the orientation of the checkerboard. Once calibration is performed at certain locations, we interpolate to obtain the light

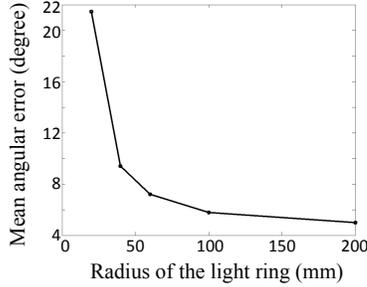


Figure 11. Mean angular error as a function of the baseline.

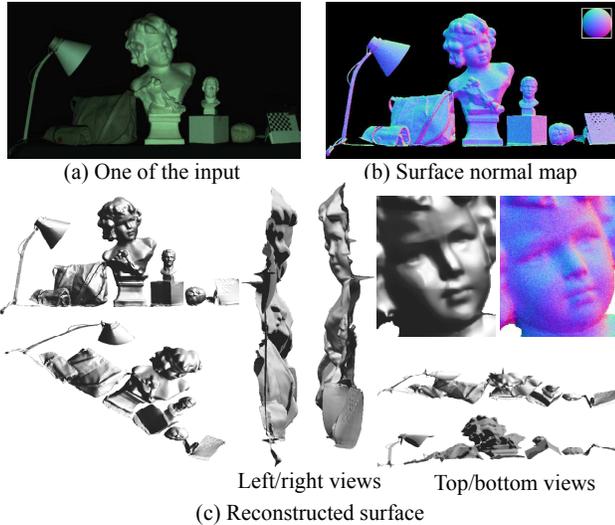


Figure 12. Surface normal map and reconstructed surface of a sculpture scene of $1m \times 0.6m$.

matrix at the remaining locations.

Results. We used the same diffuse checkerboard pattern for light calibration to quantitatively test our device. We rotated the checkerboard at 20 poses and computed the difference between the recovered surface normal and the ground truth (obtained by the use of the Camera Calibration Toolbox for Matlab [1]). The change of mean angular error with radius of the light ring is shown in Fig. 11. The error when baseline is $40mm$ is smaller than 10° .

We show the surface normals and reconstructed surfaces of real scenes captured by our device in Fig. 12. The radius of the light ring was set to $r = 40mm$. In spite of this small baseline, it is encouraging to see reliable normal estimates. Finally, in Fig. 13, we obtain surface normal estimates of a scene while varying the radius of light ring. We observe that for a radius of $20mm$, the surface normal estimates are noisy. Beyond a radius of $40mm$, the gains due to increasing baseline seem to be minimal. These results confirm the theoretical predictions outlined in Figs. 7 and 8.

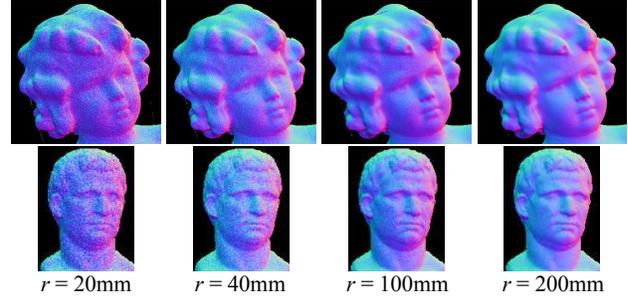


Figure 13. Surface normals of a real scene for varying baseline. Note that there is only marginal improvement in the quality of the surface normals beyond $r = 40mm$.

6. Conclusions

Photometric stereo with a small SWaP can be immensely useful. In this paper, we provide the theoretical scaffolding for understanding the dependence of estimation error as a function of various system parameters including the radius of the light ring, the number of light sources, measurement noise level as well as the dependence of the error on the location of the scene point. The systems we consider are the photometric duals of micro baseline stereo [13, 21, 15, 3]. However, to the best of our knowledge, we are the first to address the small baseline problem in the context of photometric stereo. We believe that our analysis will be useful in the design of compact and mobile photometric stereo.

Acknowledgments. J. W. and A. C. S. were supported, in part, by the NSF grant CCF-1117939. Y. M. was partially supported by JSPS KAKENHI Grant Numbers 26540085a and 15H06345. B. S. was partially supported by the Singapore MOE Academic Research Fund MOE2013-T2-1-159 and the SUTD Digital Manufacturing and Design (DManD) Centre which is supported by the Singapore National Research Foundation.

References

- [1] J.-Y. Bouguet. Camera calibration toolbox for matlab. http://www.vision.caltech.edu/bouquetj/calib_doc/, 2007. 7, 8
- [2] J. J. Clark. Active photometric stereo. In *IEEE Conf. Computer Vision and Pattern Recognition*, 1992. 2
- [3] J. Delon and B. Rougé. Small baseline stereovision. *J. Mathematical Imaging and Vision*, 28(3):209–223, 2007. 8
- [4] O. Drbohlav and M. Chantler. On optimal light configurations in photometric stereo. In *IEEE Intl. Conf. Computer Vision*, 2005. 1, 3
- [5] R. Feris, R. Raskar, L. Chen, K.-H. Tan, and M. Turk. Multiflash stereopsis: Depth-edge-preserving stereo with small baseline illumination. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 30(1):147–159, 2008. 3

- [6] R. Gardner and A. Carnesale. The solid angle subtended at a point by a circular disk. *Nuclear Instruments and Methods*, 73(2):228–230, 1969. 4
- [7] C. Hernández, G. Vogiatzis, G. J. Brostow, B. Stenger, and R. Cipolla. Non-rigid photometric stereo with colored lights. In *IEEE Intl. Conf. Computer Vision*, 2007. 2
- [8] T. Higo, Y. Matsushita, N. Joshi, and K. Ikeuchi. A hand-held photometric stereo camera for 3-d modeling. In *IEEE Intl. Conf. Computer Vision*, 2009. 2, 3
- [9] K. Horn. Shape from shading: A method for obtaining the shape of a smooth opaque object from one view. *Ph.D. thesis, Massachusetts Institute of Technology*, 1970. 2
- [10] Intel. Realsense. <http://goo.gl/DrLGHM>, 2014. 1
- [11] M. K. Johnson, F. Cole, A. Raj, and E. H. Adelson. Micro-geometry capture using an elastomeric sensor. *ACM Transactions on Graphics*, 30(4):46–53, 2011. 3
- [12] A. Jones, G. Fyffe, X. Yu, W.-C. Ma, J. Busch, R. Ichikari, M. Bolas, and P. Debevec. Head-mounted photometric stereo for performance capture. In *IEEE Conf. Visual Media Production*, 2011. 3
- [13] N. Joshi and C. L. Zitnick. Micro-baseline stereo. Technical Report MSR-TR-2014-73, May 2014. 8
- [14] Microsoft. Kinect. <http://www.microsoft.com/en-us/kinectforwindows/>, 2014. 1
- [15] G. Morgan, J. G. Liu, and H. Yan. Sub-pixel stereo-matching for dem generation from narrow baseline stereo imagery. In *IEEE Intl. Geoscience and Remote Sensing Symposium*, 2008. 8
- [16] R. Raskar, K.-H. Tan, R. Feris, J. Yu, and M. Turk. Non-photorealistic camera: depth edge detection and stylized rendering using multi-flash imaging. *ACM Transactions on Graphics*, 23(3):679–688, 2004. 3
- [17] R. Shiradkar, P. Tan, and S. H. Ong. Auto-calibrating photometric stereo using ring light constraints. *Machine vision and applications*, 25(3):801–809, 2014. 3
- [18] W. Silver. Determining shape and reflectance using multiple images. *Master's thesis, MIT*, 1980. 1, 2
- [19] D. Vlasic, P. Peers, I. Baran, P. Debevec, J. Popović, S. Rusinkiewicz, and W. Matusik. Dynamic shape capture using multi-view photometric stereo. *ACM Transactions on Graphics*, 28(5):174–184, 2009. 2
- [20] R. Woodham. Photometric method for determining surface orientation from multiple images. *Optical engineering*, 19(1):139–144, 1980. 1, 2
- [21] F. Yu and D. Gallup. 3d reconstruction from accidental motion. In *IEEE Conf. Computer Vision and Pattern Recognition*, 2014. 8
- [22] Z. Zhou and P. Tan. Ring-light photometric stereo. In *European Conference on Computer Vision*. 2010. 3
- [23] Z. Zhou, Z. Wu, and P. Tan. Multi-view photometric stereo with spatially varying isotropic materials. In *IEEE Conf. Computer Vision and Pattern Recognition*, 2013. 3