Near-Light Photometric Stereo using Circularly Placed Point Light Sources

Chao Liu
chaoliu1@cs.cmu.edu

Srinivasa G. Narasimhan
srinivas@cs.cmu.edu

Artur W. Dubrawski
awd@cs.cmu.edu

Abstract

Most photometric stereo approaches assume distant or directional lighting and orthographic imaging. However, when the source is divergent and is near the object and the camera is projective, the image intensity of a Lambertian object is a non-linear function of both the unknown surface normals and the unknown distances of the source to the surface points. The resulting non-linear optimization is non-convex and highly sensitive to the initial guess. In this paper, we propose a two-stage near-light photometric stereo method using circularly placed point light sources (commonly seen in recent consumer imaging devices like NESTcam, Amazon Cloudcam, etc). We represent the scene using a 3D mesh and directly optimize the vertices of the mesh. This reduces the complexity of the relationship between surface normals and depths in the image formation model. In the first stage, we optimize the vertex positions using the differential images induced by small changes in light source position. This procedure yields a strong initial guess for the second stage that refines the estimations using the raw captured images. We propose an accurate calibration approach to estimate the positions of the sources. Our approach performs better on simulations and on real Lambertian scenes with complex shapes than the state-of-the-art method with near-field lighting.

1. Introduction

Recovering surface shape is important for a wide range of applications such as robot manipulation, cultural heritage digitization and skin surface analysis *etc*. Photometric stereo methods use shading cues from images captured with varying illumination to recover surface shape. Traditional photometric stereo methods assume that light sources are distant, thus the lighting directions for all scene points are parallel. This is true for light sources such as the sun or for indoor lights placed far away from small objects. Under the distant light assumption, we are able to linearly solve the surface normal given the image intensities with calibrated or uncalibrated light source directions.

However, the distant light source assumption fails when the object-to-light distance becomes small. In the near-light setting, the image intensity depends non-linearly on the 3D location and normal of the scene point as well as the 3D light source position. Furthermore, for objects close to a perspective camera, the widely assumed orthographic projection model also fails. In this case, the relation between the surface normal and the depth of a scene point, which are often defined in image coordinates (at each pixel), becomes more complex when back-projected to 3D. Thus, solving for the 3D shape of an object that is illuminated by near light sources and that is captured by a projective camera is a highly non-linear and non-convex problem. As a result, photometric reconstruction often fails without strong initial guesses, as shown in Fig.1 (d).

In this paper, we present a near-light photometric stereo algorithm with circularly-placed point light sources and a perspective camera. This algorithm includes three novel contributions. First, we model the scene as a 3D triangulated mesh whose vertices correspond to the observed pixels, and directly optimize the positions of the vertices. The key advantage of this representation is that the vertex normals can be simply computed using adjacent triangular faces of the 3D mesh. The alternative of representing surface normals as numerical derivatives of depths in image coordinates (e.g. $N = (z_x, z_y, 1)$) results in unnecessary complexity when back-projected to 3D. Second, we split the algorithm into a two-stage process. In the first stage, we solve photometric stereo using the differential images captured by changing the light source position in a small amount along a circular path. We show that the analytical form of how the vertex position is related to measured differential intensity is less complex and results in reliable estimates in most parts of the object. In the second stage, these vertex positions are refined using the original image formation model applied to the raw captured images.

The above algorithm is still sensitive to errors in calibration. The light source positions in 3D are often obtained using multiple specular spheres of known radii and locations[21, 14]. But the 3D positions of these specular spheres are hard to measure accurately, resulting in poor localization of the sources for the proposed algorithm. Thus, as a third contribution, we present a simple calibration ap-

(a) Object profile   (b) Camera and Light Sources   (c) One input image (d) Method in [22]   (e) Our method
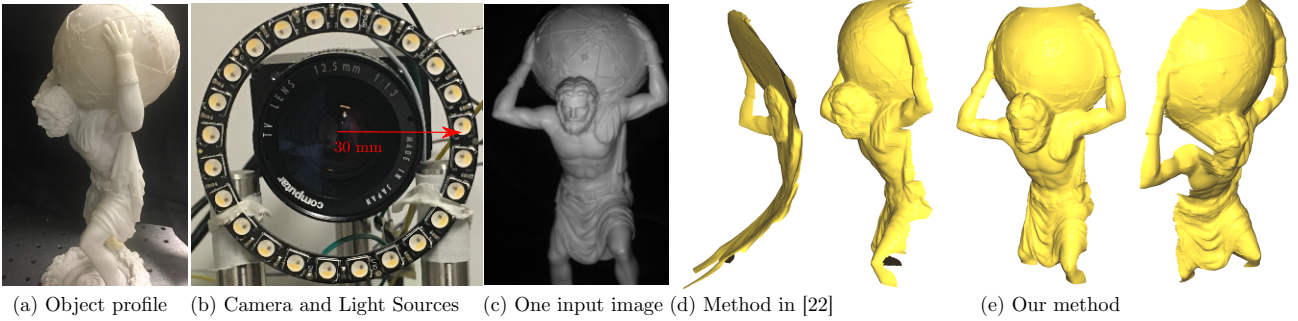
Figure 1. Near-Light Photometric Stereo using LED light sources placed in a planar circular ring centered around the camera lens (a) The profile for the reconstructed object; (b) Our imaging setup with a 30 mm radius ring of 24 LEDs; (c) One of the 24 input images; (d) Due to non-convexity of the near-light photometric stereo problem, reconstruction using [22] fails for depth initialization far away from the true values. (d) Reconstruction using our two-stage method that directly optimizes a 3D mesh. The depths and surface normals are estimated as described in Sec.3, with initializations estimated using the method in Sec.4.

proach that uses a flat panel display to estimate the source positions. The flat panel display specularly reflects light, and unlike spheres, can be calibrated precisely for position and orientation using camera calibration methods [13].

Together, the three contributions lead to effective performance on both synthetic and real scenes with complex shapes placed at various distances from the source/camera. Our method outperforms previous state-of-the-art in near-lighting photometric stereo, where the optimization suffers from poor initial guess (see Fig.1). Our approach also outperforms distant-light photometric stereo methods, even when the distance of the object is several times (5X-10X) than the radius of the LED ring. As a side effect of using differential images, our method tends to perform better in the presence of diffuse inter-reflections (but we make no claim on eliminating these effects). Our system is portable and can be implemented using a small off-the-shelf LED ring. Thus, we believe this work is timely enabling photometric 3D reconstruction on consumer imaging devices like Cloud-cam, Nest-cam that increasingly use small LED rings for imaging nearby indoor and outdoor scenes.

2. Related Work

Photometric Stereo with Distant Light Sources: Since the first formulation of the Photometric Stereo problem in [26] for shape reconstruction, there have been numerous works on improving and generalizing the method by taking into account different aspects during image formation, camera calibration and light source variations. In [18], the shape is recovered using inter-reflections by modeling the inter-reflections with form factor. For translucent objects, subsurface scattering has been taken into account in [8] and [11]. The volumetric scattering for the under-water imaging scenario is modeled in [17]. A good survey and benchmark

dataset can be found in [23]. Light intensity calibration error has been considered in [6]. The solution space for Photometric Stereo and the ambiguity in the recovered shape have been discussed in [3, 4].

Photometric Stereo with Near-field Sources: The parallel illumination direction assumption fails when the light source is close to the object. In this case, the light source is modeled as a point light source (quadratic fall-off) and the illumination direction depends on the 3D location of the scene point. In [22], a variational method is proposed to solve the inverse problem. In [19] and [15], the near-light photometric stereo is solved without calibrating the light source. In [25], a thorough analysis for reconstruction error in the near-light setup is performed. All these approaches are highly sensitive to initial guesses and do not use differential lighting based approach proposed in this work. In [27] and [24], the near-light photometric constraint is added to the multi-view scene reconstruction pipeline from images captured with different camera views. In contrast, our work is based on a single perspective view.

Photometric Stereo using Differential Lighting: In [16, 9, 12, 10], gradient illumination implemented either with a light dome or a ring of LEDs is used for surface reconstruction of human faces. In [30], the ring LED setup is used as an additional constraint during reconstruction. In [5], the differential motion of light source in the 1D circular trajectory is used for reconstructing surface with unknown BRDF. But in all these works, the sources are assumed to be distant. Most closely related is the work of [7] where the scene depth is solved directly using images captured with small near-field point light source motion. However, in order to solve for the scene depth, 3 motion directions are needed for each light source position. In contrast, the trajectory of the point light sources in our case is just a 1D curve, *i.e.* a planar circular ring, with 2 degrees of freedom
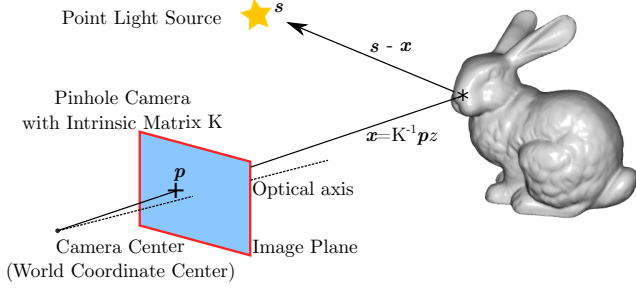
Figure 2. Image formation with near-field light source and a projective camera. For each pixel $\mathbf{p}$, there is only one variable depth $z(\mathbf{p})$. Given K, the position of the point in 3D is $\mathbf{x} = K^{-1}\mathbf{p}z$. Its surface normal $\mathbf{n}(\mathbf{p})$ is determined by $z(\mathbf{p})$ and the depths of its surrounding points as describe in Sec.3.

less for the light source motion compared to [7].

Our method is closely related to Xie *et al.* [28], where the near-light photometric stereo problem is formulated in terms of mesh deformations. Our method is different from [28] in three aspects: (1) We model the scene with 3D triangular meshes and optimize the depths of the vertices directly. The surface normals are determined directly from the 3D positions of the vertices. So there is only one variable for each pixel. Xie *et al.* [28] represent the scene as a rectangular mesh with both the surface normals and depths as variables (similar to many other previous works). So there are three variables for each pixel, making it harder to optimize. This is redundant since for a mesh representation, normals are completely determined by vertice positions; (2) The method in [28] assumes orthographic camera model. Thanks to the scene representation, our method works for perspective cameras; (3) Because we determine the surface normal from the vertice depths, our method does not rely on mesh deformation to get the depths from surface normals, as in [28]. This leads to robustness to depth discontinuities for our method.

3. Near-Light Photometric Stereo on a 3D Mesh

In this section, we describe the image formation model for near-light photometric stereo of a Lambertian object illuminated by point light sources and captured by a perspective camera. Without loss of generality, we set the origin for the world coordinate frame to be the center of the camera, as shown in Fig. 2. The albedo and the surface normal for a scene point are denoted by ρ and $\mathbf{n}$. The point source is at location $\mathbf{s}$. Then radiance R of a scene point at $\mathbf{x}$ is:

$$R = \rho s_e \frac{\mathbf{n}^T(\mathbf{s} - \mathbf{x})}{|\mathbf{s} - \mathbf{x}|^3} = \tilde{\rho}\frac{\mathbf{n}^T(\mathbf{s} - \mathbf{x})}{|\mathbf{s} - \mathbf{x}|^3}, \qquad (1)$$

where s_e is the light source intensity; $\tilde{\rho} = \rho s_e$ is the scaled albedo. The cubic in the denominator accounts for the normalization for the incident light vector and the quadratic fall-off of light intensity in the point light source model.

The scene point is imaged by a camera with intrinsic matrix K. We define the homogeneous image coordinate for the point $\mathbf{x}$ projected on the image plane to be $\mathbf{p}$. For a scene point with depth z, the image coordinate $\mathbf{p}$ and the world coordinate $\mathbf{x}$ are related by back-projection:

$$\mathbf{x} = K^{-1}\mathbf{p}z \qquad (2)$$

Combining Eq.1 and Eq.2, the image intensity $I(\mathbf{p}; z, \mathbf{n})$ for the scene point $\mathbf{x}$ can be written as:

$$I(\mathbf{p}; z, \mathbf{n}) = \tilde{\rho}\frac{\max\{\mathbf{n}^T(\mathbf{s} - K^{-1}\mathbf{p}z), 0\}}{|\mathbf{s} - K^{-1}\mathbf{p}z|^3}, \qquad (3)$$

where, attached shadow is modeled using the $\max$ operator.

It is hard to optimize for surface normals and depths as separate unknowns. Thus, we need to exploit their relationship. However, representing surface normals as numerical derivatives of depths in image coordinates (e.g. $N = (z_x, z_y, 1)$) results in unnecessary complexity when back-projected to 3D. Instead, we represent the scene as a 3D mesh with triangular faces F whose vertices V are defined for all image pixels. We then use the vertex normal for calculating the image intensity in Eq. 3. This process is illustrated in Fig. 3(a). The 3D location of the vertex v_i is $\mathbf{x}_i$ and its 2D imaged location is $\mathbf{p}(v_i)$. Given the depth $z(v_i)$ for vertex v_i, $\mathbf{x}(v_i)$ is given by $\mathbf{x}(v_i) = K^{-1}\mathbf{p}(v_i)z(v_i)$. An adjacent face f consists of v_i and two other vertices v_j and v_k. The edges connecting v_i to v_j and v_k are $\mathbf{e}_{ij} = \mathbf{x}_j - \mathbf{x}_i$ and $\mathbf{e}_{ik} = \mathbf{x}_k - \mathbf{x}_i$ respectively. Then, the unnormalized vertex normal for v_i is defined as:

$$\hat{\mathbf{n}}(v_i) = \frac{\sum\limits_{f \in \mathcal{N}_f(i)} a(f)\mathbf{n}(f)}{\sum\limits_{f \in \mathcal{N}_f(i)} a(f)} = \frac{\sum\limits_{f \in \mathcal{N}_f(i)} [\mathbf{e}_{ij}]_\times \mathbf{e}_{ik}}{\sum\limits_{f \in \mathcal{N}_f(i)} a(f)} \qquad (4)$$

where $\mathcal{N}_f(i)$ are the neighboring faces that include vertex v_i; $\mathbf{n}(f)$ and $a(f)$ are the normal and area for face f. The vertex normal $\mathbf{n}(v_i)$ is obtained by normalizing $\hat{\mathbf{n}}(v_i)$:

$$\mathbf{n}(v_i) = \hat{\mathbf{n}}(v_i)/|\hat{\mathbf{n}}(v_i)| \qquad (5)$$

We solve for the depths of the vertices by combining image formation model in Eq.3, Eq.4 and Eq.5. However, there remains a depth ambiguity for vertex v_i if we use the normal definition in Eq.4. Consider the cases shown in Fig. 3(b) and (c). The faces around vertex v_i are related by rotations around the vertex normal $\mathbf{n}_v$. Because the face areas are the same, the horizontal components of the surface normals of the neighboring faces are canceled out in the weighted sum in Eq.4. So the vertex normals in two
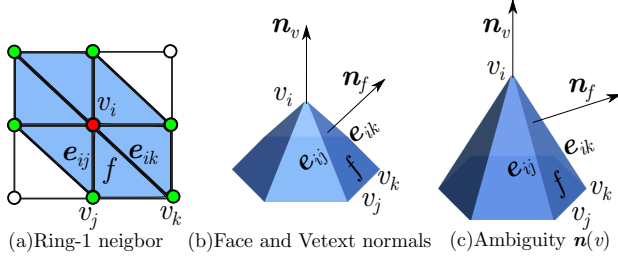
(a)Ring-1 neigbor (b)Face and Vetext normals (c)Ambiguity $\boldsymbol{n}(v)$

Figure 3. Local geometry around vertex v_i. (a) The surface normal at vertex v_i, colored in red, depends on the depths of its ring-1 neighborhoods, colored in green; (b) We define the surface normals to be the vertex normal $\mathbf{n}(v_i)$, which depends on the face normals of meshes sharing v_i (c) The original vertex normal definition leads to ambiguity in the depth estimation for v_i. We deal with this ambiguity using the method presented in Sec.3.

cases both point perpendicularly upwards, even though the locations for vertex v_i and the face normals are different.

We solve this ambiguity by changing the order of shading in Eq.3 and computing the weighted averaged in Eq.4: we first compute the image intensities for each face around vertex v_i, then compute weighted average of intensities to get the intensity for vertex v_i. More formally, the image intensity $I(v_i; z(v_i))$ for vertex v_i is re-written as:

$$I(v_i; z(v_i)) = \tilde{\rho} \frac{\sum_{f \in \mathcal{N}_f(i)} a(f) \mathcal{S}\left(\mathbf{s}, \hat{K}, z(v_i), f\right)}{|\mathbf{s} - \hat{K}(v_i) z(v_i)|^3 \sum_{f \in \mathcal{N}_f(i)} a(f)}, \quad (6)$$

with $\hat{K} = K^{-1} \mathbf{p}(v_i)$ and the shading operator defined as:

$$\mathcal{S}\left(\mathbf{s}, \hat{K}, z(v_i), f\right) = \max\{\mathbf{n}^T(f)(\mathbf{s} - \hat{K} z(v_i)), 0\}$$

3.1. Objective Function

We estimate the depth values for all vertices by minimizing the difference between the modeled intensities I in Eq.6 and the measured image intensities $\tilde{I}$:

$$\begin{aligned} \min_z \quad & \sum_{v_i \in V} \left(\tilde{I}(v_i) - I\left(v_i; z(v_i)\right) \right)^2 + \lambda_I E_s(z(v_i)) \\ \text{with} \quad & E_s\left(z(v_i)\right) = \sum_{v_j \in \mathcal{N}_v(i)} (z(v_i) - z(v_j))^2 \end{aligned} \quad (7)$$

where $\mathcal{N}_v(i)$ is the set of vertices in the Ring-1 neighborhood of v_i. The albedo $\tilde{\rho}$ is solved analytically using Eq.6.

The energy function for the optimization problem defined in Eq.7 has a numerous local minima due to the cubic term in the denominator in $I(v_i; z(v_i))$ defined in Eq.6. So, good initializations of the depth values at vertices is crucial, as validated using experiment shown in Fig.4. The face of a
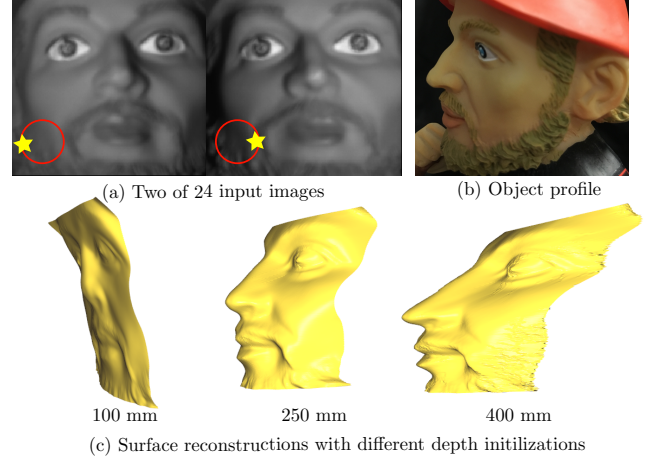

(a) Two of 24 input images (b) Object profile

100 mm   250 mm   400 mm

(c) Surface reconstructions with different depth initilizations

Figure 4. Sensitivity of the optimization of depths of vertices to initial guesses with planes in different depths. The light sources are a 30 mm radius ring of 24 LEDs centered around the camera. The object is placed 200 mm away from the camera. (a) Two out of 24 input images, with the LED positions at the lower left corner. (b) The profile of the reconstructed area for reference. (d) The profiles of the surfaces reconstructed with different initial depths.

toy is reconstructed using the L-BFGS minimizer for Eq. 7 with different initial depth values. We use a ring of 24 LEDs centered around the camera. The radius of the LED ring is 30 mm and the object is placed around 200 mm away from the camera. More details about the implementation are included in Sec.6. As shown in Fig.4, if the initialization of depths is far away from the true values, we get either over-flattened or stretched results due to the local minima.

4. Near-Light Photometric Stereo with Differential Circular Source Motion

In order to get a good initial guess for the optimization problem in Eq.7, we put forward to use the differential images induced by small change of light source position. The illumination and imaging geometry is shown in Fig. 5. For notation simplicity, for LED at position $\mathbf{s}$ and with index t, we denote the differential source motion as $\mathbf{s}_t$ and the corresponding differential image intensity as I_t, which means the differential values $w.r.t$ to the LED index. By differentiating the image formation model in Eq.1, we get the analytical form for the differential image intensity I_t:

$$\begin{aligned} I_t &= \frac{\partial I}{\partial \mathbf{s}} \mathbf{s}_t \\ &= \tilde{\rho} \frac{\mathbf{n}^T \mathbf{s}_t}{|\mathbf{s} - \mathbf{x}|^3} - 3\mathbf{n}^T(\mathbf{s} - \mathbf{x}) \frac{\tilde{\rho}(\mathbf{s} - \mathbf{x})^T \mathbf{s}_t}{|\mathbf{s} - \mathbf{x}|^5} \\ &= \tilde{\rho} \frac{\mathbf{n}^T \mathbf{s}_t}{|\mathbf{s} - \mathbf{x}|^3} - 3I \frac{(\mathbf{s} - \mathbf{x})^T \mathbf{s}_t}{|\mathbf{s} - \mathbf{x}|^2} \end{aligned} \quad (8)$$
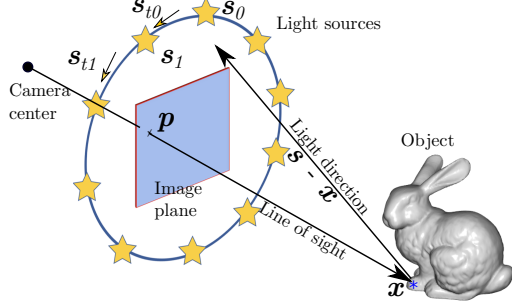
Figure 5. The geometry of differential change of light source positions. The light sources are densely mounted on a planar circle centered around the camera.

which can be simplified by writing the first term as a function of the image intensity I:

$$I_t = I \frac{\mathbf{n}^T \mathbf{s}_t}{\mathbf{n}^T (\mathbf{s} - \mathbf{x})} - 3I \frac{(\mathbf{s} - \mathbf{x})^T \mathbf{s}_t}{|\mathbf{s} - \mathbf{x}|^2} \qquad (9)$$

Intuitively, the first term is the contribution of the change of source direction; the second term is due to the change of distance between the light source and the scene point.

For light sources mounted on a plane parallel to the image plane, there are two special cases where the second term including the inverse squared distance in Eq.9 becomes small and can be ignored.

The first case occurs when the angle between the light direction $\mathbf{s} - \mathbf{x}$ and the light source motion vector $\mathbf{s}_t$ is large such that $(\mathbf{s} - \mathbf{x})^T \mathbf{s}_t \approx 0$. This happens when the object is placed far away from the light source such that $\mathbf{s} - \mathbf{x}$ is perpendicular to the light source plane spanned by $\mathbf{s}_t$. However, for large distance between the light source and scene point, both the captured image intensity I and the differential image intensity I_t would be too small with low SNR.

The second case where we can ignore the second term in Eq.9 is when the source motion trajectory is circular. Here, the source motion direction $\mathbf{s}_t$ and position $\mathbf{s}$ are perpendicular, thus $\mathbf{s}^T \mathbf{s}_t = 0$. So, the differential image intensity I_t in Eq.9 becomes:

$$I_t = I \frac{\mathbf{n}^T \mathbf{s}_t}{\mathbf{n}^T (\mathbf{s} - \mathbf{x})} + 3I \frac{\mathbf{x}^T \mathbf{s}_t}{|\mathbf{s} - \mathbf{x}|^2} \qquad (10)$$

For a camera with a traditional field of view, the angle between the direction of line of sight and the normal of image plane is small. Since the plane spanned by $\mathbf{s}_t$ is parallel to the image plane, $\mathbf{x}^T \mathbf{s}_t$ in the second term in Eq.10 becomes small. The term $\mathbf{x}^T \mathbf{s}_t$ is further attenuated by the squared distance between the point and the light source $|\mathbf{s} - \mathbf{x}|^2$. So we can assume that the second term in Eq.10 for small light source motions can be ignored, at least for the purposes of estimating our initial guess.

Thus, the differential image intensity and the measured image intensity are related by:

$$I \mathbf{n}^T \mathbf{s}_t - I_t \mathbf{n}^T (\mathbf{s} - \mathbf{x}) = 0 \qquad (11)$$

This is similar to the differential image term in [5] for the Lambertian case. The difference is that we can solve for the depths using Eq.11 alone rather than obtaining a constraint for the surface normal as in [5].

More specifically, given the measured image $\tilde{I}$ and differential image $\tilde{I}_t$, the depths can be estimated by:

$$\min_z \sum_{v_i \in V} E_{It}(v_i; z(v_i)) + \lambda_{I_t} E_s(z(v_i)) \qquad (12)$$

with

$$\begin{aligned} E_{It}(v_i; z(v_i)) &= \left(\hat{\mathbf{n}}(v_i)^T \left(\tilde{I}(v_i) \mathbf{s}_t - \tilde{I}_t(v_i)(\mathbf{s} - \mathbf{x}(v_i)) \right) \right)^2 \\ \mathbf{x}(v_i) &= K^{-1} \mathbf{p}(v_i) z(v_i) \end{aligned}$$

Note that the energy function in Eq.12 is independent from the inverse squared distance. Thus the energy function is less non-linear than the one in Eq.7. In addition, since the function is independent from the albedo, given the measured image intensities $\tilde{I}$ and differential image $\tilde{I}_t$, we can estimate the depths without knowing the surface albedo.

5. Complete Algorithm and Calibration

We use the optimized depth in Eq.12 as the initial values in Eq.7 where we estimate the depth values with raw image intensity I. To initialize the optimization problem in Eq.12, we use a line search for the depth of each vertex. Given the candidate depth value $z_c(v_i)$, we can solve for the corresponding candidate surface normal $\mathbf{n}_c$ using Eq.1. Then we validate the candidate depth value $z_c(v_i)$ and $\mathbf{n}_c$ using the differential image I_t with Eq.9. For each vertex, we choose the candidate depth value that minimizes the difference between the measured and modeled I_t, to be the initial depth. The complete near-light photometric stereo with circular placed LEDs is summarized in Algorithm.1.

5.1. Localizing Light Sources

It is important to calibrate the 3D light source positions accurately since we use the first-order derivative of the source positions. The calibration error introduced by traditional calibration methods using one or multiple chrome spheres will fail our algorithm, since the precise 3D location and projected radius of the sphere in the image plane required by these methods are difficult to measure or calibrate automatically. Instead, we propose a light source position calibration method using a flat specular display: First, we display the checkerboard pattern on a planar glossy display such as the monitor of a Macbook and capture one image

Algorithm 1 Near-light Photometric Stereo with Circular Placed LEDs

1: Given images I, differential images I_t, Camera Intrinsic Matrix K, Light Source Positions S and Light Source motion vectors S_t;
2: Initialize the depths with line search for each vertex.
3: Estimate the depths z_{I_t} using Eq. 12 ;
4: Initialize the albedo $\tilde{\rho}_{I_t}$ given z_{I_t};
5: Initialize: $z_I^{(0)} = z_{I_t}$, $\tilde{\rho}^{(0)} = \tilde{\rho}_{I_t}$, $k = 0$
6: **for** k $\in \{1, \cdots, MaxIter\}$ **do**
7:      Get $z_I^{(k)}$ using Eq.7, with $z_I^{(k-1)}$, $\tilde{\rho}^{(k-1)}$ as the initials
8:      Given $z_I^{(k)}$, solve for $\tilde{\rho}^{(k)}$ using Eq.6
9: **end for**
10: return $z_{opt} = z_I^{(k)}$, $\tilde{\rho}_{opt} = \tilde{\rho}^{(k)}$
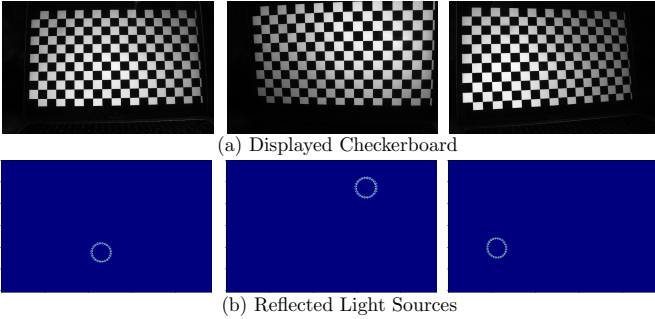

(a) Displayed Checkerboard

(b) Reflected Light Sources

Figure 6. We calibrate the camera and the point light source positions using a planar glossy display (Macbook monitor). (a)Images captured with the display turned on and light source turned off, from which we estimate the camera intrinsic matrix and plane parameters. (b)Superposition of the images captured when the display is turned off and the light sources are turned on.

for each setup of the plane, as shown in Fig.6(a). For each plane orientation, we turn off the display and turn on the LEDs sequentially and capture one image for each LED.

For each plane setup, we can get the plane parameters (plane orientation and distance from the origin) from the well-established camera calibration process [13]. Given the camera intrinsic matrix and plane parameters , for each light source reflection, we then estimate the light position by ray-tracing and triangulating for the centers of highlights in the light source reflection images shown in Fig.6(b).

We evaluate the performance of our calibration procedure and compare it with the method using chrome spheres [21]. To demonstrate the sensitivity of the calculated light source locations *w.r.t.* the sphere center estimation, we evaluate the performance with/without using the ground truth 3D position of the sphere center. In the case where the ground truth 3D position is unknown, we use the ellipse detector in [20] to get the ellipse parameters. Then we approximate the ellipse to be round circle with center same
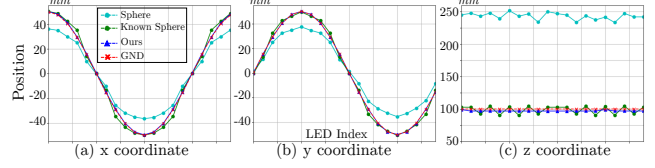

(a) x coordinate   (b) y coordinate   (c) z coordinate

Figure 7. The xyz profiles of the locations for all point light sources. Cyan - sphere chrome based method; The ellipse parameters are estimated using [20]; Green - sphere chrome based method; The ground truth location of the sphere center is given. Blue - Our method; Red - Ground truth.

as the ellipse center and radius as the mean of axes lengths of the ellipse. Given the physical dimension of the sphere and the camera focal length, we can get the depth and the location of the sphere center as in [21]. The comparison results in simulation are shown in Fig.7. As shown by the cyan curve in Fig.7, inaccurate estimations for the 3D locations of sphere centers lead to large errors, especially in the z direction. By contrast, our method is much more accurate.

6. Experiments

6.1. Implementation Details

We implement the differential motion of the light source with an LED ring with 24 LEDs and 30 mm in radius. The reconstructed objects are placed 300 - 400 mm away from the camera and light sources. We use the Prosilica GT1930c camera manufactured by Allied Vision to capture the images. Each image is captured with .1 second exposure time with one LED turned on. The algorithm is implemented in Python and C++, with the Ceres-Solver [1] for optimization. For the energy functions defined in Eq.7 and Eq.12, we set the weights for the smoothness term E_s to be $\lambda_I = .1$ and $\lambda_{I_t} = .01$ respectively. For faster convergence, we perform the optimizations in multiple scales where the results from lower resolution are used as the initializations for the higher resolution. The running time for 968×608 image resolution is about 5 min using on a desktop with Intel Core-i7 5940 CPU and 64 GB RAM memory size. We will release the implementation upon publishing.

6.2. Simulations Results

We test our algorithm with different imaging setups with synthesized images. To validate the effectiveness of the initialization using differential images, we place the reconstructed surface at a plane with 900 mm depth, facing towards the camera. We reconstruct the surface with and without the first stage of the proposed method. For the compared method with no initializations, we set initial depth to be 200 mm. We run this comparison for multiple image settings where the LED ring radius ranges from 20 mm to
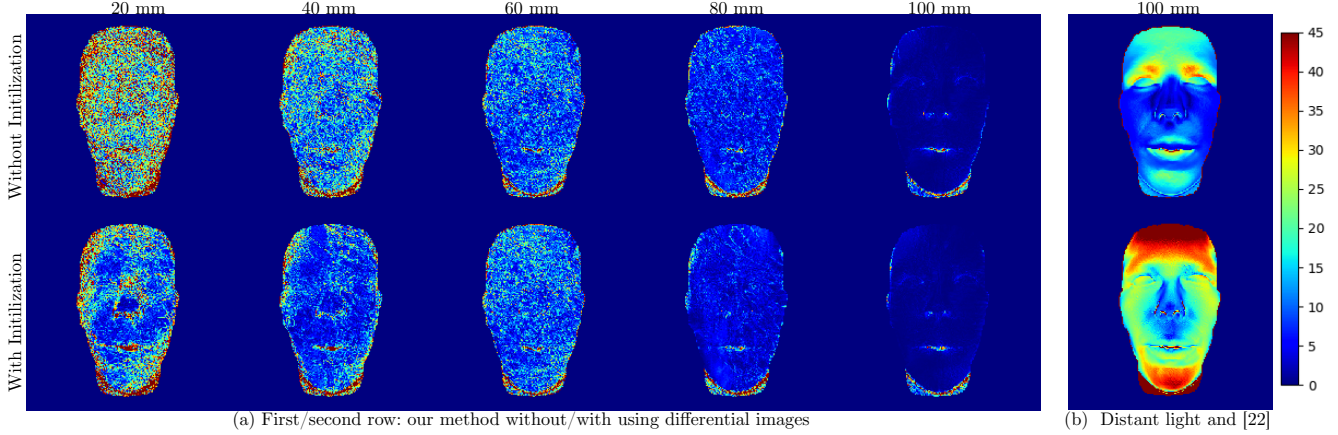
(a) First/second row: our method without/with using differential images   (b) Distant light and [22]

Figure 8. Ablation study for using the differential image I_t in Eq.12 with different LED ring radii. The mean depth for the object is 900 mm. For radius $= 100\ mm$, we compare our method with photometric stereo with distant light assumption, and the method in [22]. (a) Our method without and with the depth estimation using Eq.12 to initialize the depths for the optimization in Eq.7. Initialization using the differential images in Eq.12 helps to improve the reconstruction, especially for smaller light source baselines. For both setups, the error decreases with larger LED baselines. In our real experiment setup, we use an LED ring with a radius of 30 mm. (b) Error maps using distant light assumption (first row) and method in [22] (second row). Both methods result in large errors even for the largest LED baseline.

Figure 9. The input images with the light sources on the LED ring turned on sequentially. The radius of the LED ring is 30 mm and the object is placed around 400 mm away. The image pair shown in the same column corresponds to LED pairs on the opposite sides of the ring. As shown, even for the largest LED baselines (the image pair shown in one column), the difference between images is still small.

(a) Reconstructed Object   (b)Distant light with $z_{\mathrm{init}} = 200mm$ (c)[22] with $z_{\mathrm{init}} = 200\ mm$ (d)Distant light with initialized z_{lt} (e) [22] with initialized z_{lt} (f) Our method
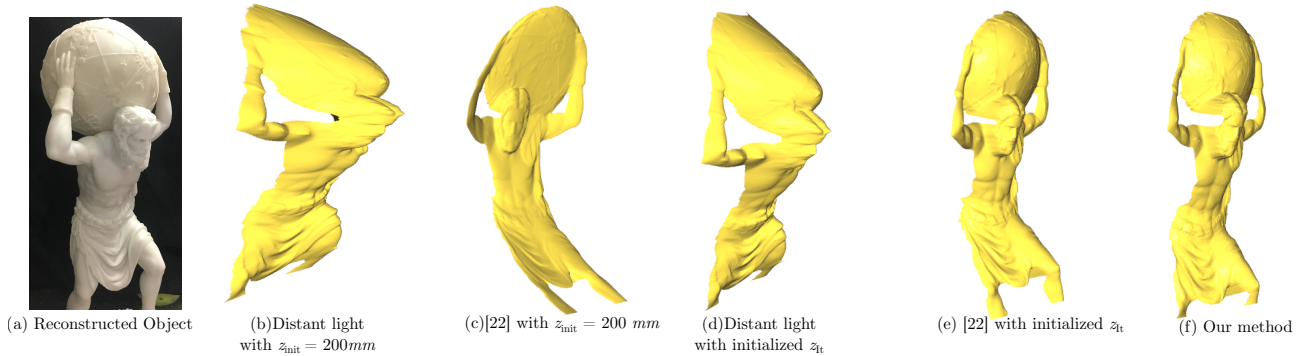
Figure 10. We show the effectiveness of the first-stage of results by showing the reconstruction for the atlas statue using different methods with/without initialization using z_{I_t}. The performance for both compared methods increases by using z_{I_t}. (a)Object profile; (b)Reconstruction with distant light source assumption; (c)Reconstruction using method in [22]; (d)(c)Reconstruction with the comparison methods, with z_{I_t} as the initialization; (f) Our method.

100 mm. Then we measure the angles between the estimated and ground truth surface normals to quantify the performance. As shown in Fig.8, the initialization using the differential images helps to improve the reconstruction, especially for smaller light source baselines. For larger LED ring radius, the performance of the method proposed in Sec.3 without initialization is comparable to the method in Sec.5 with depth initialization. This might be because there are less local minima for the energy function in Eq.7 for larger light source baselines. With the same 200 mm depth initialization, both photometric stereo using distance light source assumption and the method in [22] induce large errors even for the largest light source baseline.

We evaluate our method for 6, 10, 14 and 18 LEDs with the same scene setup and LED ring dimension. The mean surface normal error is 10.42, 3.15, 2.63 and 2.56 degrees respectively. Note that the error sharply drops as we increase the LEDs from 6 to 10. This shows that the small light source baseline makes the approximation in Sec.4 valid even for fewer LEDs.

6.3. Real World Results

We apply our algorithm on images captured with LEDs on a ring with 30 mm in diameter. One sequence of capture images is shown in Fig.9 for a bust statue placed around 400 mm away from the camera. Due to the small ratio between the LED ring radius and the object-to-camera distance, the difference between images is very small even for the pair of LEDs with the largest baselines. Despite the small difference, our method still performs well as shown in Fig.11 where both the large depth variation between the left and right shoulders, and the fine grained structures in the frontal clothes are reconstructed. Note that for small baseline near-light photometric stereo, the optimization for depths is more easily trapped in the local minima due to the fact that the changes in the intensities are small. Thus using just the image intensity I is likely to generate degraded reconstructions if the initialization is not good, as shown both in Fig.4 with our image formation in Eq.2 , and in Fig.1(d). One extreme case would be that the baseline for the light source is zero. In this case the problem becomes the highly ill-posed shape from shading problem and can be solved only with prior-knowledge about the shape geometries [2, 29]. In our case, for near-light photometric stereo with small light source baselines, the initializations using I_t helps the optimization in Eq.7 process to avoid poor initializations and keep it from getting trapped in the surrounding local minima in the first stage.

To further validate the effectiveness of the first-stage of our method, we apply both traditional photometric stereo under distant lighting assumption and the method in [22] with the estimation results z_{It} as the initial depth values. For distant-light photometric stereo, we use those initialized

depth values to get the lighting directions for all points; for the method in [22], we use z_{It} as its initial depth guess for optimization. As shown in Fig.10, by using z_{It}, the performance for both compared methods increases. Note in Fig.10(e), the reconstructed for the right leg of atlas is in accurate by bending forward, even though we have initialized the depth estimation process with I_{It} for this case.

We apply our method on other objects with different scene geometries. Results are shown in Fig.11. The first two columns of Fig.11 are two input images taken with lights on the opposite sides of the LED ring turned on. Although the image difference is small, our method is able to recover both the overall shape with enough depth variations, such as the shape of face, and fine-grained details, such as the logo on the tennis shoe. More results are shown in the supplementary material.

For small light source position changes, we assume that the global component does not change much. As a result, we cancel out the global illumination component by subtracting two images captured with close light sources during estimating I_t. Based on this observation, we can further refine the reconstruction results by adding another step where the analytical form of I_t in Eq.9 is used for optimizing depths, with z_{opt} in Algorithm 1 as the initial values. We test this idea for reconstructing the object surface with large concavity such as the bowl shown in Fig.12. As shown, we get more robustness against the global illumination component by using the differential images I_t.

7. Limitations

The Lambertian assumption in our method fails when the surface includes specular reflection component, as can be seen in the human face reconstruction example in the last row of Fig.11. This leads to high-frequency artifacts such as the spike on the reconstructed nose. Another limitation of our method is that even though using the differential images leads to more robustness to the global light component as shown in Fig.12, the global component is not fully modeled and removed during reconstruction. So the reconstruction error in the presence of global component is still observable. One future direction is to include both the BRDF model and global illumination term into our problem formation.

8. Conclusion

In this paper, we put forward a two-stage near-light photometric stereo algorithm with circularly placed point light sources and a pinhole camera. In the first-stage, we optimize the scene depth using the differential images captured by moving the light source slightly. We show in the paper that the surface reconstruction becomes less non-linear by using the differential images. In the second stage, we refine the estimations using the raw captured images. We validate
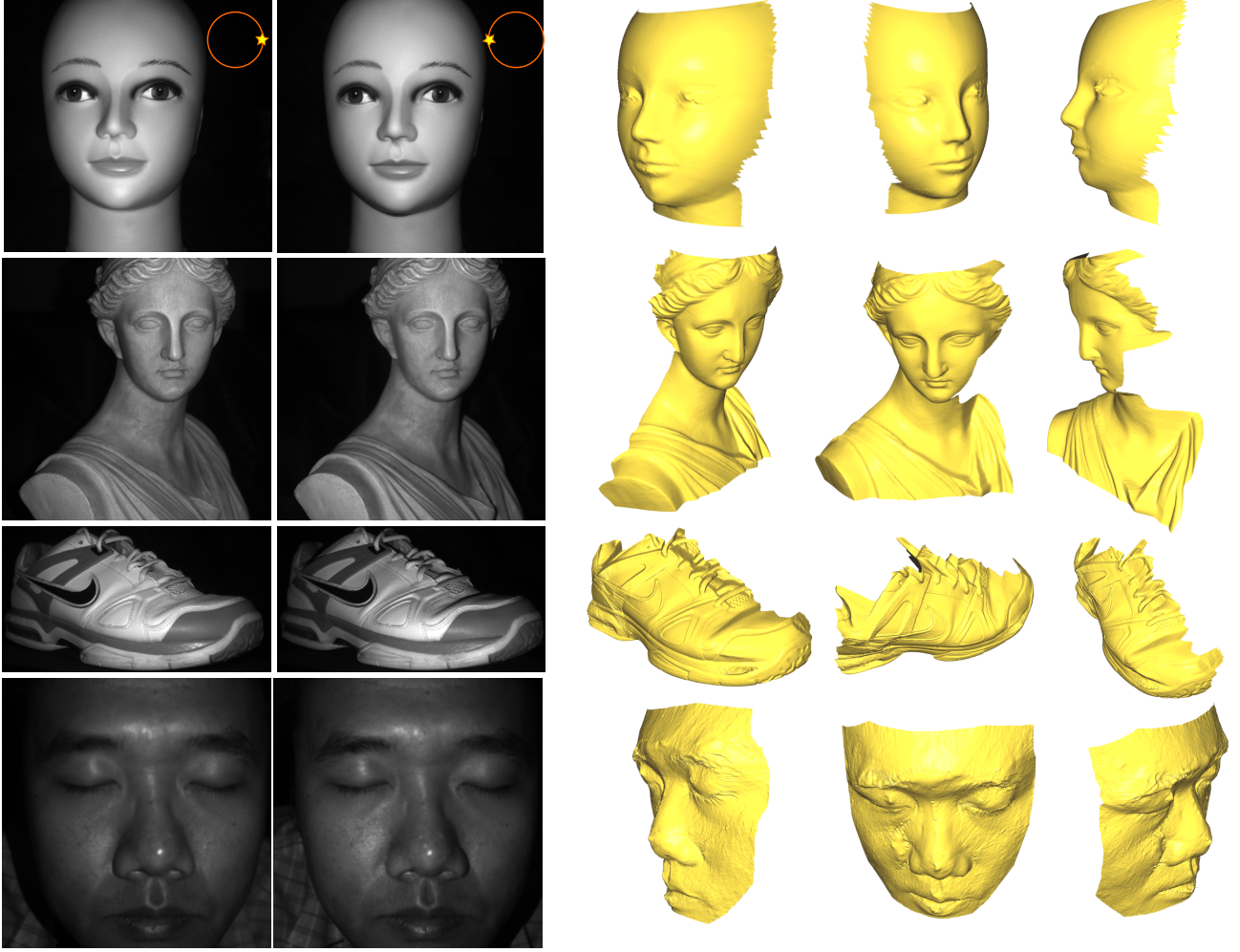
Figure 11. Inputs and reconstruction results using the proposed method. From left to right: Two of 24 input images taken with lights on the opposite sides of the LED ring turned on. Reconstructed meshes viewed from different views.

(a) One input image   (b) Depth estimation using I   (c) Depth estimation using I and It   (d) Reference depth map using structured light   (e) Depth profiles
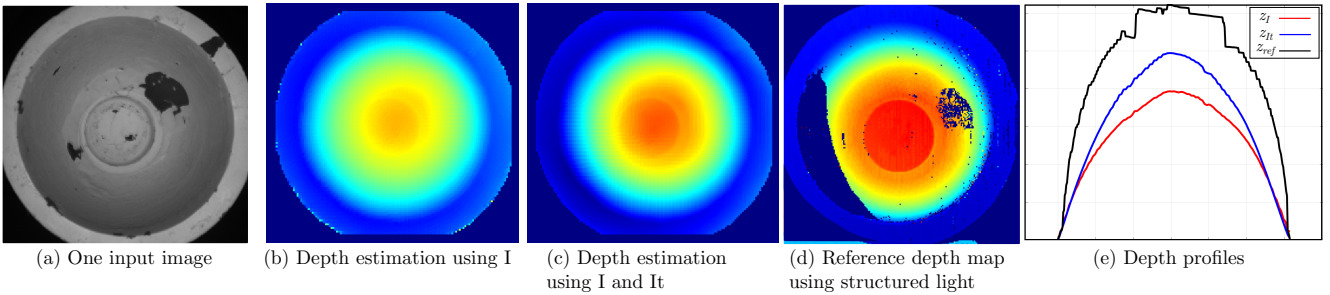
Figure 12. We reconstruct the inner surface of a concave bowl. We get more robustness against the inter-reflection during reconstruction with the differential images I_t. From left to right: (a) one of the input images; (b) estimated depth map using the raw images I only; (c) estimated depth map using both I and differential images I_t; (d) reference depth map estimated using structured light with global-direct light separation; (e) 1D depth profiles for different methods (red: using I only; blue: using I and I_t; black: reference).

that our method is able to get good reconstruction results even with small baseline point light sources such as a low-cost LED ring. One future direction is to consider cases with general BRDFs and global illumination.

9. Acknowledgments

References

[1] S. Agarwal, K. Mierle, and Others. Ceres solver. `http://ceres-solver.org`.

[2] J. T. Barron and J. Malik. Shape, illumination, and reflectance from shading. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(8):1670–1687, 2015.

[3] P. N. Belhumeur and D. J. Kriegman. What Is the Set of Images of an Object Under All Possible Illumination Conditions ? *International Journal*, 260(28):245–260, 1998.

[4] P. N. Belhumeur, D. J. Kriegman, and A. L. Yuille. The Bas-Relief Ambiguity. *International Journal of Computer Vision*, 35(1):33–44, 1999.

[5] M. Chandraker, J. Bai, and R. Ramamoorthi. On differential photometric reconstruction for unknown, isotropic BRDFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(12):2941–2955, 2013.

[6] D. Cho, Y. Matsushita, Y.-W. Tai, and I.-S. Kweon. Photometric Stereo Under Non-uniform Light Intensities and Exposures. *ECCV*, pages 170–186, 2016.

[7] J. Clark. Active Photometric Stereo. In *CVPR*, 1992.

[8] B. Dong, K. D. Moore, W. Zhang, and P. Peers. Scattering parameters and surface normals from homogeneous translucent materials using photometric stereo. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2299–2306, 2014.

[9] A. Ghosh, G. Fyffe, B. Tunwattanapong, J. Busch, X. Yu, and P. Debevec. Multiview face capture using polarized spherical gradient illumination. *Proceedings of the 2011 SIGGRAPH Asia Conference on - SA '11*, 30(6):1, 2011.

[10] P. Graham, B. Tunwattanapong, J. Busch, X. Yu, A. Jones, P. Debevec, and A. Ghosh. Measurement-based synthesis of facial microgeometry. *Computer Graphics Forum*, 32(2 PART3):335–344, 2013.

[11] C. Inoshita, Y. Mukaigawa, Y. Matsushita, and Y. Yagi. Surface Normal Deconvolution: Photometric Stereo for Optically Thick Translucent Objects. In *ECCV*, pages 346–359. 2014.

[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. *Conference for Visual Media Production*, pages 158–164, 2011.

[13] J.Y.Bouguet. MATLAB calibration toolbox. `http://www.vision.caltech.edu/bouguetj/calib_doc/`.

[14] J. Liao, B. Buchholz, J. M. Thiery, P. Bauszat, and E. Eisemann. Indoor Scene Reconstruction Using Near-Light Photometric Stereo. *TIP*, 26(3):1089–1101, 2016.

[15] J. Liao, B. Buchholz, J. M. Thiery, P. Bauszat, and E. Eisemann. Indoor Scene Reconstruction Using Near-Light Photometric Stereo. *IEEE Transactions on Image Processing*, 26(3):1089–1101, 2017.

[16] 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 *Eurographics Symposium on Rendering (EGSR)*, 2007.

[17] Z. Murez, T. Treibitz, R. Ramamoorthi, and D. Kriegman. Photometric stereo in a scattering medium. *Proceedings of the IEEE International Conference on Computer Vision*, (October):3415–3423, 2015.

[18] S. K. Nayar, K. Ikeuchi, and T. Kanade. Shape from interreflections. *International Journal of Computer Vision*, 6(3):173–195, 1991.

[19] T. Papadhimitri and P. Favaro. Uncalibrated Near-Light Photometric Stereo. In *BMVC*, pages 1–12, 2014.

[20] V. Pătrăucean, P. Gurdjos, and R. G. von Gioi. A parameterless line segment and elliptical arc detector with enhanced ellipse fitting. In *ECCV 2012*.

[21] M. W. Powell, S. Sarkar, and D. Goldgof. A simple strategy for calibrating the geometry of light sources. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(9):1022–1027, 2001.

[22] Y. Quéau, B. Durix, T. Wu, D. Cremers, F. Lauze, and J.-D. Durou. Led-based photometric stereo: Modeling, calibration and numerical solution. *Journal of Mathematical Imaging and Vision*, Sep 2017.

[23] B. Shi, Z. Wu, Z. Mo, D. Duan, S.-K. Yeung, and P. Tan. A Benchmark Dataset and Evaluation for Non-Lambertian and Uncalibrated Photometric Stereo. *CVPR*, pages 3707–3716, 2016.

[24] Tomoaki Higo, Yasuyuki Matsushita, Neel Joshi, and Katsushi Ikeuchi. A hand-held photometric stereo camera for 3-D modeling. *ICCV*, (Iccv):1234–1241, 2009.

[25] J. Wang, Y. Matsushita, B. Shi, and A. C. Sankaranarayanan. Photometric stereo with small angular variations. In *ICCV*, volume 2015 Inter, pages 3478–3486, 2015.

[26] R. J. Woodham. Photometric method for determining surface orientation from multiple images, 1980.

[27] C. Wu, S. G. Narasimhan, and B. Jaramaz. A multi-image shape-from-shading framework for near-lighting perspective endoscopes. *International Journal of Computer Vision*, 86(2-3):211–228, 2010.

[28] W. Xie, C. Dai, and C. C. L. Wang. Photometric stereo with near point lighting: A solution by mesh deformation. In *CVPR*, volume 07-12-June, pages 4585–4593, 2015.

[29] Y. Xiong, A. Chakrabarti, R. Basri, S. J. Gortler, D. W. Jacobs, and T. Zickler. From shading to local shape. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(1):67–79, 2015.

[30] Z. Zhou and P. Tan. Ring-light Photometric Stereo. In *ECCV*, 2010.