

# Photometric Stereo in the Wild

Chun-Ho Hung<sup>#</sup> Tai-Pang Wu<sup>†</sup> Yasuyuki Matsushita<sup>§</sup> Li Xu<sup>‡</sup> Jiaya Jia<sup>†</sup> Chi-Keung Tang<sup>◇</sup>  
<sup>#</sup> Quora, Inc. <sup>†</sup> The Chinese University of Hong Kong <sup>‡</sup> Lenovo R&T  
<sup>§</sup> Microsoft Research Asia <sup>◇</sup> Hong Kong University of Science and Technology

## Abstract

Conventional photometric stereo requires to capture images or videos in a dark room to obstruct complex environment light as much as possible. This paper presents a new method that capitalizes on environment light to avail geometry reconstruction, thus bringing photometric stereo to the wild, such as an outdoor scene, with uncontrolled lighting. We do not make restrictive assumption, and only use simple capture equipments, which include a mirror sphere and a video camera. Qualitative and quantitative experiments indicate the potential and practicality of our system to generalize existing frameworks.

## 1. Introduction

Digitalizing 3D objects from still images is one of the core problems in this field. To recover a highly detailed surface, photometric stereo (PS) is known to be one effective method. Since Woodham [25] introduced the concept, research has been done to improve the estimation quality, computation robustness with various reflectance models.

One critical problem that arguably hinders PS from being widely employed is the stringent lighting requirement. Lighting directions and their intensities need to be known to uniquely reconstruct the target surface. In calibrated photometric stereo, the number of light sources as well as their corresponding lighting direction needs to be known in order for PS to work. This condition makes data acquisition generally be performed in a dark room to eliminate environment lighting and inter-reflection.

As a result, PS will fail when the target object receives an unexpected light emitted or reflected from the environment, as exemplified in Fig. 1. In the laboratory scene in (a), even after all lights are switched off, there is still ambient light reflected or scattered from the wall and furniture when illuminated by a spotlight. Two indoor scenes in (b)-(c) contain different levels of light inter-reflection. Light passing through windows further complicates the situation. The last example shown in (d) is an open area, which makes it even more difficult to control the lighting condition.

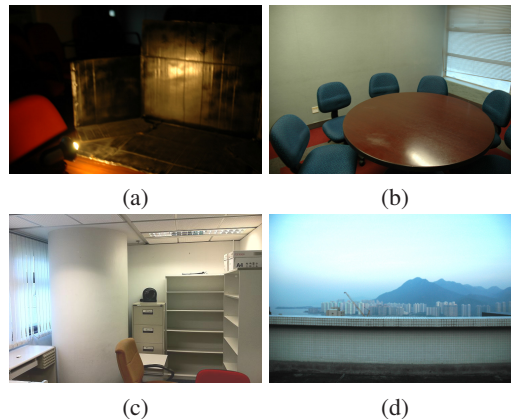


Figure 1. Typical scenes that are difficult for photometric stereo.

To generalize PS, instead of physically controlling the environment, which sometimes is impossible, a better way is to naturally include the captured light in estimation such that ubiquitous ambient light will not be decisively harmful, but on the contrary help produce decent results.

In this paper, we propose a simple but effective technique to make use of complex environment light to work at our advantage, allowing the image capturing process in PS be carried out in an uncontrolled environment including outdoor. Our method does not assume only sunlight exists. Instead, we use a low-cost mirror sphere configuration to accurately and quickly *capture* complex environment lights. We propose a simple setup for PS that captures complex lights for normal and depth estimation.

**Characteristics of our method** Our system has the following main advantages compared to traditional ones.

- (a) We simply capture the whole environment lighting map for each image, regardless the lighting is direct/dominant illumination or otherwise;
- (b) Traditionally, a direct light source needs to be moved to provide a sufficient number of different configurations for producing necessary photometric parallax for solving the system of equations. In our method, any alteration in the environment, which is not restricted to dominant light source, can be used to produce photometric parallax.

**Assumptions** The assumptions that apply to our work are those also adopted in traditional PS.

- (i) The class of objects that we handle depends on the descriptiveness of the chosen analytical model. The Lambertian model is discussed in detail in this paper.
- (ii) All of the environment light sources are distant.

These conditions are not strict. As will be shown, for (i), we found that the Lambertian model in the environment lighting configuration can satisfactorily handle many objects with our outlier-rejection framework. For (ii), empirically we produce good results inside a small room where lights are close to the objects that arguably violates the distant light source assumption.

## 2. Related Work

In this section, we first briefly review photometric stereo methods. As our focus in this paper is to propose a new setting for general light capturing, we will discuss previous light inference/estimation steps.

**Classical PS Methods** Given a static Lambertian scene as well as a number of distant light sources and corresponding lighting directions, Woodham [25] showed that at least three images are required to recover the surface orientation for each pixel. In [6], four images were used where shadow and peculiarity are identified as outliers. To handle more objects, the Torrance-Sparrow model [23] was proposed to deal with non-Lambertian surfaces. A hybrid reflectance model was introduced in [15] to extract local surface orientation along with reflectance.

**Modern PS Methods** Compared with these classical methods, modern PS techniques have a higher degree of freedom for both the surface/reflectance representation and experimental setting.

Representative work includes the example based methods [11, 9] to transfer normals from reference or virtual objects to the target one; the glossy surface method using cast shadow for normal estimation [5]; bivariate approximation of the isotropic reflectance function to model spatially varying reflectance [3]; sub-pixel surface normal estimation method [24] using Expectation Maximization (EM); method to tackle the anisotropic reflectance effect [13]; the dense frame robust photometric stereo method [26] for casual data acquisition; colored lighting method [10] for dynamic objects; consensus method [12] to get rid of analytical models; and the self-calibration method [22] to automatically calibrate lighting information and scene radiance simultaneously.

Recently, multiview geometry was incorporated to avoid constructing wider-view 3D surfaces [14]. In [16], Nehab *et al.* proposed Multiview photometric stereo (MVPS), and

used a triangulation scanner and a photometric stereo scanner to capture 3D models. Esteban *et al.* [8] combined MVS and PS to recover complete surfaces with details. Zhou *et al.* [28] used multiview to address the spatially varying material problem. Park *et al.* [18] generated high quality 3D mesh in a parameterized space.

**Light Calibration using Spheres** The above methods either require light calibration using chrome spheres or estimating unknown lighting in a controlled environment. Many of them assume that there is a countable number of *direct* light sources.

Wu *et al.* [26] estimated lighting by identifying saturated pixels on a mirror sphere. With an additional diffuse sphere, Goldman *et al.* [9] computed relative light intensity. Park *et al.* [19] advocated the use of a weakly textured plane to calibrate near point light source. The single point-source assumption will not hold in general indoor or outdoor scenes.

There are a few methods, such as the ones described in [20, 21, 1, 27], that study PS for outdoor scenes. While these approaches address some specific problems, we propose a simple and general technique to allow data acquisition in an uncontrolled environment. Lights coming from almost all possible directions that are visible to the camera are taken into consideration in our photometric reconstruction.

Our work is essentially different from the method of [4]. In [4], no light calibration device is employed and unknown lighting and object shape are algorithmically *estimated*. Due to the high difficulty of the problem and the ambiguity of solutions, it has to assume a smooth reflectance function that can be approximated by second-order spherical harmonics and require user input to handle the generalized bas-relief (GBR) ambiguity manually. Computation time can also be an issue. In comparison, we propose a handy, simple, and general method with mirror spheres to directly *capture* lights without assuming particular environment types. The photometric stereo estimation is *automatic* and can produce high quality results in a short time.

## 3. Environment Lighting Photometric Stereo

A mirror sphere was used in computer graphics for environment capturing and photo-realistic rendering [7]. One example is shown in Fig. 2(a). We show that its proper employment in PS can enable the capture of incoming direct (*e.g.*, moving handheld spotlight, fluorescent tube) and indirect (*e.g.*, inter-reflections) illumination, thus greatly benefiting PS in general scenes.

The mirror sphere shown in Fig. 2(a) captures complex lights, even from the emissive black bodies and the environment, in addition to the major light sources that are considered in conventional PS. The rich light information allows for alleviating the restriction of only detecting major light

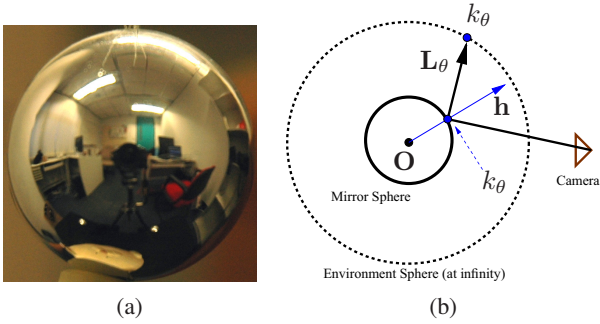


Figure 2. A mirror sphere for capturing environment lighting. (b) shows the relationship between the mirror sphere and environment sphere, where  $\mathbf{O}$  is the *ideal* location where the object is placed,  $\mathbf{h}$  is the angle bisector between the lighting direction  $\mathbf{L}_\theta$  and viewing direction.  $k_\theta$  is the incident intensity along  $\mathbf{L}_\theta$ .

through dense color sampling from the sphere. For a complex environment, all of these samples can be regarded as light sources, which make the following estimation process accurate.

Specifically, Fig. 2(b) illustrates the relationship between lighting direction  $\mathbf{L}_\theta \in \mathbb{R}^3$  and corresponding intensity  $k_\theta \in \mathbb{R}$ . It indicates that performing light capture using a mirror sphere does not need to distinguish between *direct* and *indirect* lights. They are similarly taken into consideration in our method. We show in Section 8 that with this strategy we can produce decent normal maps even if the lighting condition is unknown or uncontrollable.

Similar to the notations used in [17], the intensity  $I$  observed at a surface point can be modeled as an integration

$$I = \int_{\theta \in \Omega} f(\mathbf{L}_\theta, \mathbf{V}) k_\theta \mathbf{N}^T \mathbf{L}_\theta d\theta, \quad (1)$$

where  $f(\cdot)$  is the BRDF at the surface point,  $\Omega$  is a space containing all possible orientations,  $\mathbf{L}_\theta$  is a lighting direction defined by orientation  $\theta$ ,  $k_\theta$  is the incident intensity at  $\theta$ ,  $\mathbf{V}$  is the viewing direction at the surface point, and  $\mathbf{N}$  is the corresponding surface normal.

#### 4. Estimating Surface Normals

We adopt the Lambertian model, and show that good results can be produced for a large variety of uncontrolled scenes, which is difficult to accomplish in prior work. Eq. (1) is expressed as

$$I = \sum_{\theta \in \Omega'} \rho k_\theta \mathbf{N}^T \mathbf{L}_\theta, \quad (2)$$

where  $\rho$  is the surface albedo and  $\Omega'$  is the discrete version of  $\Omega$ . The discretized  $\Omega'$  corresponds to the sampled directions on the mirror sphere. From the camera's point of view,  $I$  is the illuminance observed at a pixel.

Note that Eq. (2) is different from the one used in [21]. We present an explicit model characterized by surface normals and lighting directions, while the method of Shen and

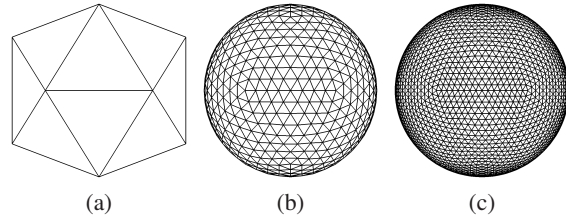


Figure 3. Icosahedron and its subdivision. (a) A 20-face polyhedron, where the vertices are evenly distributed on a 3D unit sphere that encloses and touches the polyhedron. (b) A 3-time subdivided icosahedron. (c) A 4-time subdivided icosahedron. Each subdivision is done by splitting each face into four equilateral triangles followed by reprojecting the vertices onto a unit 3D sphere.

Tan [21] is derived based on cosine kernels and coefficients in spherical harmonics to represent data obtained from internet, where recovering dense and accurate normal maps is not the main concern.

With Eq. (2), our target is to estimate  $\rho$  and  $\mathbf{N}$  from a set of captured images  $I_i$ , indexed by  $i = 1 \cdot \cdot \cdot n \mid n \geq 3$  with a fixed viewing direction. As  $\mathbf{N}^T \mathbf{L}_{\theta_i}$  cannot be negative, Eq. (2) can be more accurately expressed as

$$I_i = \sum_{\theta \in \Omega'} \rho k_{\theta_i} \max(\mathbf{N}^T \mathbf{L}_{\theta_i}, 0), \quad (3)$$

where  $k_{\theta_i}$  and  $\mathbf{L}_{\theta_i}$  are respectively the incident intensity and the lighting direction defined by  $\theta$  for image  $i$ . The  $\max(\cdot)$  operator is used to reject negative energies. Eq. (3) describes a single-pixel color formation.

Note that due to the presence of  $\max(\cdot)$ , Eq. (3) is not continuous in this form. We thus cannot simply apply existing continuous optimization methods, such as gradient descent, to solve for  $\rho$  and  $\mathbf{N}$ . Also, ignoring the  $\max(\cdot)$  operator to approximate a linear system can make the negative  $\mathbf{N}^T \mathbf{L}_{\theta_i}$  significantly affect the result. These facts feature our PS representation. Note that in the traditional configuration, with only a few light sources, rejecting negative  $\mathbf{N}^T \mathbf{L}_{\theta_i}$  is easy by removing dark regions or shadow, while in our problem, the number of negative terms is much larger. We will show later the result comparison with and without removing the negative terms (Fig. 8(d) and (c)). In what follows, we propose a two-step approach as the new solver.

##### 4.1. Surface Normal and Albedo Estimation

We propose a novel two-step method. In the first step, we estimate an initial normal and a surface albedo by selecting the most suitable candidate from a discrete set of surface normals. Afterwards, an energy function associated with Eq. (3) is formulated and is solved based on the initial normal. This strategy can quickly find the appropriate normal estimates for each pixel.

**Step 1: Discrete Approximation** Fig. 3(a) shows an image of an icosahedron. One interesting property of an ico-

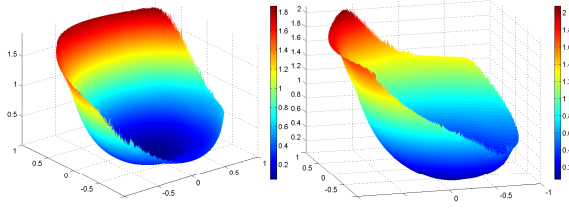


Figure 4. Each pair  $(\tilde{\mathbf{N}}, \tilde{\rho})$  yields an error  $E(\tilde{\rho})$ . The plots for two randomly selected pixels are shown, where the  $z$ -axis is  $E(\tilde{\rho})$  and the  $x$ - and  $y$ -axes represent the slant and tilt of  $\tilde{\mathbf{N}}$ . Global minima exist for all test cases, which indicates that only one pair of  $\tilde{\mathbf{N}}$  and  $\tilde{\rho}$  can produce the minimum error, which is reliably selected as initialization to step 2.

hedron is that the respective vertices are evenly distributed over a unit 3D sphere. This property also applies to differently subdivided versions of the icosahedron (Fig. 3(b)-(c)). We compute a set of possible surface orientations that are evenly distributed in all directions and denote them as the candidate normals  $\mathcal{N}$ . In experiments, we subdivide the icosahedron four times, which yields 2562 different normal candidates. Only half of them (*i.e.*, 1281 normals) are visible to the camera.

For each  $\tilde{\mathbf{N}} \in \mathcal{N}$ , we solve for  $\rho$  by minimizing the following energy

$$E(\rho) = \sum_i \|I_i - \sum_{\theta \in \Psi} \rho k_{\theta_i} \tilde{\mathbf{N}}^T \mathbf{L}_{\theta_i}\|^2, \quad (4)$$

where  $\Psi \in \mathcal{O}'$  is a space containing all lighting directions making  $\tilde{\mathbf{N}}^T \mathbf{L}_{\theta_i} \geq 0$ . Since the surface normal is fixed for each  $\tilde{\mathbf{N}}$ , we can identify  $\Psi$  easily.  $\max(\cdot)$  disappears because of the introduction of  $\Psi$ . Eq. (4) can be solved quickly by a least-square method.

After obtaining all the possible  $\tilde{\mathbf{N}}$ , we pair each  $\tilde{\mathbf{N}}$  with the associated albedo  $\tilde{\rho}$  and error  $E(\tilde{\rho})$ . The pair  $(\tilde{\mathbf{N}}^0, \tilde{\rho}^0)$  that produces the smallest error is taken as the appropriate candidate, which is the input to step 2 as initialization.

Using a number of images captured in a variable lighting environment, we find that a unique global minimum of  $E$  always exists for all examples. Since the discrete space  $\Psi$  varies w.r.t.  $\tilde{\mathbf{N}}$ , we show two examples in Fig. 4 with plotted errors. They demonstrate that global minima exist and the respective error surfaces are smooth.

**Step 2: Continuous Refinement** Since the normal space described by  $\mathcal{N}$  is discretized by a subdivided icosahedron, after obtaining the initial  $\tilde{\mathbf{N}}^0$  and  $\tilde{\rho}^0$ , we further refine them to increase the pixel-wise estimation accuracy. This is done by minimizing Eq. (4) using the continuous Levenberg-Marquardt algorithm starting from  $\tilde{\mathbf{N}}^0$  and  $\tilde{\rho}^0$ . In each iteration, when  $\tilde{\mathbf{N}}$  and  $\tilde{\rho}$  are updated, the space of  $\Psi$  has to be updated too to avoid a negative energy introduced by negative  $\tilde{\mathbf{N}}^T \mathbf{L}_{\theta_i}$ . However, in practice, we found that the result is almost the same even with fixed  $\Psi$ .

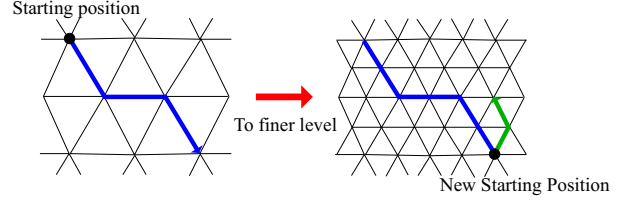


Figure 5. Coarse-to-fine estimation. Here shows the subdivided icosahedron in two levels. With the normal estimate in the coarse level, we continue the search to increase the accuracy in the finer level, which forms a path shown in the left figure. When the search stops, the estimate is further passed to the next level shown on the right until the maximum level is reached.

**Coarse-to-Fine Estimation** For a single pixel, Eq. (4) originally has to be computed  $|\mathcal{N}|$  times with  $|\Psi|$  possible lighting directions in order to examine the possible pairs of  $\tilde{\mathbf{N}}$  and  $\tilde{\rho}$ , which is computationally expensive. To speed it up, we apply a coarse-to-fine searching algorithm where estimation of  $\tilde{\mathbf{N}}^0$  starts from the original icosahedron that contains the smallest number of vertices. After the best candidate normal is found in this level (by applying step 1), we move to a finer level by subdividing the icosahedron and take the current normal estimate (represented by an icosahedron vertex) as the starting position, which is already good except that it lacks a higher degree of accuracy. We search neighboring vertices, which correspond to more normal samples, and move to the one that produces the minimum error in Eq. (4). The search is achieved by following the path minimizing Eq. (4). When the moving process stops, we go to a finer level. This repeats until the desired subdivision level, which corresponds to the degree of accuracy for the normal estimation, is reached. The process is illustrated in Fig. 5.

## 5. Data Acquisition Configuration

We now establish the correspondence between our environment lighting calibration system with those employed in a dark room where only countable major distant point light sources are used or assumed. Our major finding is that even the most complex environment lighting, if it is faithfully captured, corresponds to a *single* virtual light source in an imaginary dark room. Thus, the traditional chrome sphere light calibration can be regarded as a *special case* of our configuration.

To prove it, we express Eq. (2) as

$$I = \rho \mathbf{N}^T \sum_{\theta \in \Psi} k_{\theta} \mathbf{L}_{\theta} = \rho \mathbf{N}^T (\gamma \hat{\mathbf{L}}), \quad (5)$$

where  $\gamma = \|\sum_{\theta \in \Psi} k_{\theta} \mathbf{L}_{\theta}\|$  and  $\hat{\mathbf{L}} = \frac{1}{\gamma} \sum_{\theta \in \Psi} k_{\theta} \mathbf{L}_{\theta}$ .  $\gamma$  and  $\hat{\mathbf{L}}$  are respectively the *virtual light intensity* and *direction*. With this simplified form, it is possible to solve for  $\mathbf{N}$  given three different pairs of  $\gamma$  and  $\hat{\mathbf{L}}$ . Fig. 6 shows the



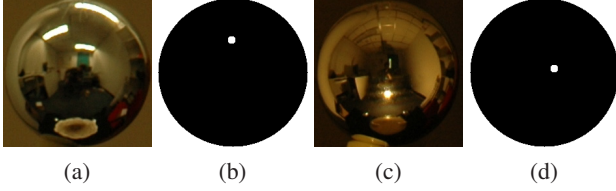


Figure 6. Environment lighting captured by a mirror sphere can be converted into a single virtual light source by Eq. (5). (a) and (c) show two mirror spheres. (b) and (d) show the virtual light sources corresponding to them in an imaginary dark room.

relationship between the image of a mirror sphere and the corresponding virtual lighting in a dark room.

This observation indicates that we actually have many ways to change light for photometric stereo in general scenes, compared to conventional method to only move the major light sources. We list a few new choices below, which can greatly expand the applicability of PS:

- (1) We can simply alter part of the environment. This can be achieved by turning on/off ceiling or floor lamps in a room, moving the whole system, or with a certain angle, or simply blocking the environment light using a cardboard.
- (2) We allow the use of a handheld spotlight even with complex environment lighting. Moving the spotlight can similarly produce photometric parallax.

Fig. 6(a) shows an indoor environment allowing lights to be turned on and off. We have also built a makeshift system using a shopping cart for transporting the camera, the sphere, and the object to different places for photometric stereo reconstruction where lighting can hardly be controlled. The second acquisition method was exemplified in Fig. 6(c), using a handheld spotlight.

In theory, a minimum number of three images is sufficient for solving Eq. (4) with photometric parallax. More inputs generally help increase the accuracy and robustness of the system. Seven or more images are used empirically when noticeable cast shadow arises. The cast shadow issue will be addressed in Section 7.

## 6. Issues

We discuss possible issues related to the use of mirror spheres for environment light calibration here.

**Saturated Pixels** The mirror image can be saturated for some pixels when highlight exists. This problem can be addressed using high-dynamic range (HDR) imaging [7], which linearizes the camera response curve and makes the pixel intensity proportional to the ground truth radiance.

**Sampling of  $\Omega'$**  The image of a mirror sphere is the 2D projection of a 3D sphere. It spatially suffers from non-uniform sampling of radiance – that is, the center of the

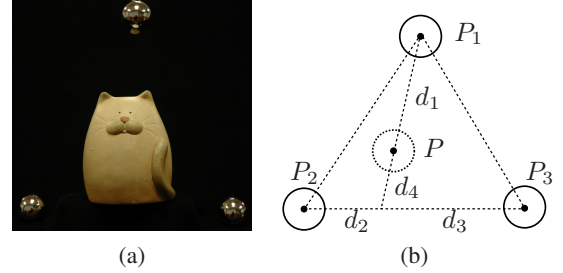


Figure 7. (a) The placement of the mirror spheres. (b) An illustration of mirror image interpolation.

sphere image has a higher sampling rate than the boundary. So we should not use the pixel color directly from the image. Instead, the subdivided icosahedron are employed, which are aligned with the mirror sphere such that the pixels that coincide with respective vertices are considered for sampling. We subdivide the icosahedron three times, which yields 642 sampling points.

Given any sample, we can determine its surface normal and take it as the angle bisector to estimate the corresponding  $L_\theta$  using the law of mirror reflection. Before sampling, the mirror sphere image is low-passed by a Gaussian filter to avoid aliasing. Its standard deviation is set to be proportional to the edge length of the subdivided icosahedron.

**Placement of Mirror Spheres** In practice, the object to be captured is not at the same position as the mirror sphere. So, similar to all other light calibration configurations for PS using chrome/diffuse spheres, estimation errors can be caused by the bias of lighting sampled from the sphere. We propose capturing additional information from multiple mirror spheres, which are placed at different locations near the object, to alleviate the problem.

We use three mirror spheres on a plane approximately parallel to the image plane. One sphere is above the object and the other two are respectively located at the lower left and right corners. One example is shown in Fig. 7(a). We blend the three mirror sphere images using triangular interpolation (as illustrated in Fig. 7(b)):

$$P = \frac{d_4}{d_1 + d_4} P_1 + \frac{d_1}{d_1 + d_4} \left( \frac{d_3}{d_2 + d_3} P_2 + \frac{d_2}{d_2 + d_3} P_3 \right),$$

where  $P$  is the resultant mirror sphere image corresponding to a virtual sphere,  $P_1$ - $P_3$  are the three mirror sphere images, and  $d_1$ - $d_4$  are lengths illustrated in Fig. 7(b), coarsely measuring the distances to the center of the object.

## 7. Outlier Removal

With a non-convex shape or non-Lambertian materials of the captured object, cast shadow and highlight are sometimes unavoidable as shown in Fig. 8(a), which could affect the accuracy of geometry estimation. For one pixel, if we can reliably find more than three input images, in which the

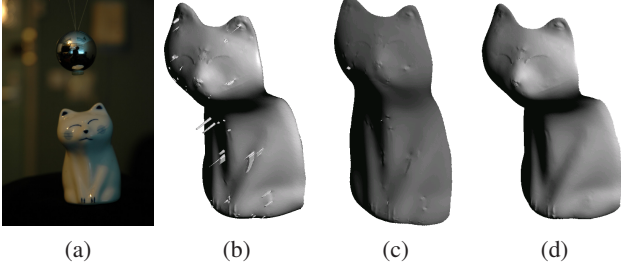


Figure 8. Shadow and highlight outliers. One of the 12 input images is shown in (a). The surface results without and with outlier removal are shown in (b) and (d) respectively. If negative  $\mathbf{N}^T \mathbf{L}_{\theta_i}$  terms are included in computation (described in Section 4), even with outlier rejection, the result is problematic, as shown in (c).

pixel is not in shadow or highlight, the normal and albedo estimates ( $\tilde{\mathbf{N}}^0$  and  $\tilde{\rho}^0$ ) are accurate. To identify a few of these frames, we propose a RANSAC scheme described below. Note that we are now free from negative  $\tilde{\mathbf{N}}_0^T \mathbf{L}_{\theta_i}$  terms and it is safe to apply RANSAC.

1. Randomly choose 3 images from the input and compute a normal-albedo pair ( $\tilde{\mathbf{N}}^0, \tilde{\rho}^0$ ), using the algorithm described in Section 4.
2. For each image  $i$ , compute the squared error  $E'_i$  with respect to the analytical model  $E'_i = \|I_i - \sum_{\theta \in \Psi} \tilde{\rho}_0 k_{\theta_i} \tilde{\mathbf{N}}_0^T \mathbf{L}_{\theta_i}\|^2$ .
3. Compute the set of inlier images:  $T = \{i \mid E'_i \leq \epsilon_i\}$ .
4. Re-estimate the normal-albedo pair ( $\tilde{\mathbf{N}}_T, \tilde{\rho}_T$ ) using all images in  $T$  by the algorithm described in Section 4.
5. Perform step 2 again after replacing ( $\tilde{\mathbf{N}}_0, \tilde{\rho}_0$ ) by ( $\tilde{\mathbf{N}}_T, \tilde{\rho}_T$ ). Also update the set of inliers in step 3.
6. Repeat steps 1-5  $K$  times. Keep track of ( $\tilde{\mathbf{N}}_T, \tilde{\rho}_T$ ) that has the largest set of inliers.

This procedure samples different image triples to identify outliers. The intensity equations are formed by the normal and albedo reported as inliers, given that in the input images highlight and shadow are not the majority. The final per-pixel estimates are chosen as those that correspond to the largest-size inlier set based on a credible error measure. This scheme is effective to remove problematic pixels in surface estimation.

To deal with brightness variation, the threshold for each image  $\epsilon_i$  is set to be proportional to the average of the environment light intensity. Empirically, if the number of input images is small, we can replace the random process by exhaustion of all image triples. Fig. 8(c) shows that our outlier rejection notably improves the surface estimates.

## 8. Experimental Results

We evaluate our technique in a variety of challenging environments, in which traditional photometric stereo is hard

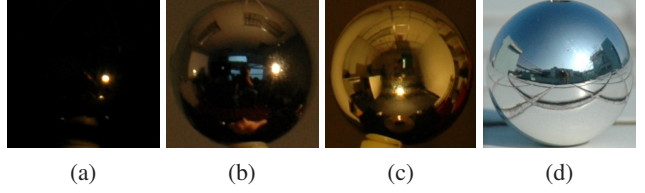


Figure 9. Environments for data capturing in our experiments. (a) Ideal dark room. (b) Mirror in Fig. 1(b), containing low inter-reflection. (c) Mirror in Fig. 1(c) with high inter-reflection. (d) Challenging outdoor scene for Fig. 1(d).

Example	Resolution	# of images	One-layer time	C2F time
<b>Syn. Case</b>	101×101	9	49.7s	0.16s
<b>KitCat</b>	357×442	14	1499.1s	7.02s
<b>HumanDoll</b>	299×714	10	253.1s	4.84s
<b>Kitten</b>	415×666	12	538.2s	8.33s
<b>WineBottle</b>	548×870	8	2048.6s	7.23s
<b>Vase</b>	268×576	12	285.9s	5.45s

Table 1. Running time for the examples shown in this paper. It is reported on a desktop computer with a Core2Duo CPU 2.79GHz and 2GB memory. “C2F” stands for “coarse-to-fine”.

to apply. Fig. 9 shows the mirror images. The scenes contain different levels of inter-reflection and lighting. Table 1 lists the running time. The color coded normal is rendered as  $(R = (u_x + 1)/2, G = (u_y + 1)/2, B = u_z)$  from surface normal  $(u_x, u_y, u_z)^T$ . The final surfaces are constructed using the method of [2].

**Toy Example** We synthesized a set of images for sanity check purpose. Fig. 10 shows an example, where the input 9 images are rendered with environment lighting. With the ground truth surface normal in (c), which is a hemisphere, we compare the results produced using traditional PS methods [26] without considering environment lighting and our new approach. They are respectively shown in images (d)-(f) and images (g)-(i). It is noticeable that ignoring the environment lighting flattens the resulting surfaces. The average angular errors (AAE) are 28.85 and 0.39 degrees respectively, differing hugely.

Environment	Surface Normal AAE
(a) Dark-room	3.234815
(b) Low inter-reflection	5.621666
(c) High inter-reflection	8.134300

Table 2. Average angular errors (AAE) of the estimated normal maps, corresponding to the results shown in Fig. 11(d)-(f).

**Different Environment Evaluation** We put *KitCat* in different challenging environments as shown in Fig. 1(a)-(c) for surface reconstruction. The object in Fig. 11(a) contains

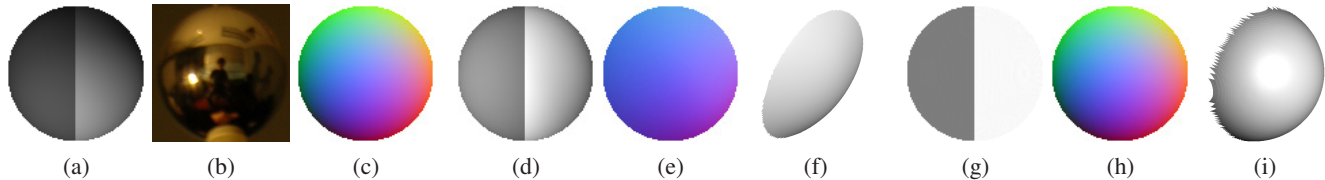


Figure 10. *Synthetic Case* – (a) One out of 9 input images. (b) The corresponding mirror sphere. (c) Ground truth normal map (color coded). (d)-(f) Albedo and normal maps, and the reconstructed surface by only finding the major light sources. (g)-(i) Our results, close to the ground truth.

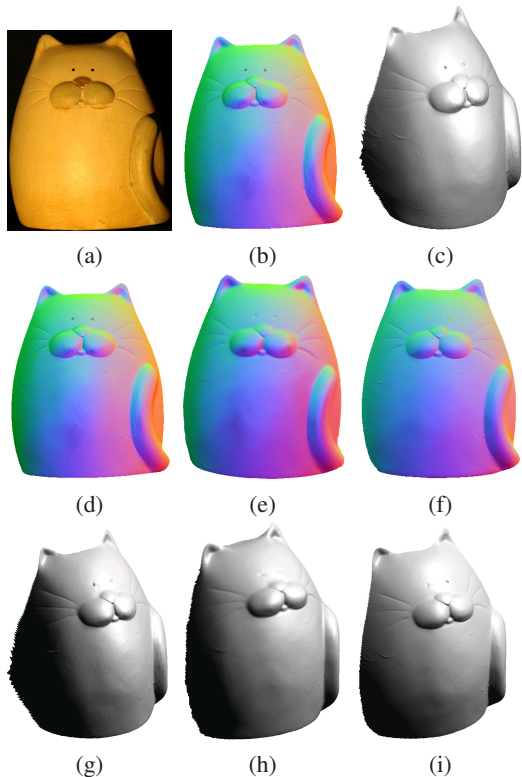


Figure 11. *KitCat* – (a) An input image. (b) Ground truth normal map generated in a dark room. (c) The reconstructed surface from (b). (d)-(f) Color coded normal maps generated by our method with the environments respectively shown in Fig. 1(a)-(c). (g)-(i) Corresponding surfaces.

glossy reflection. In all of the environments, we test both using a handheld lamp to illuminate the object and switching on and off different ceiling fluorescent tubes to produce photometric parallax. Our quantitative evaluation shown in Table 2 by comparing to the ground truth result obtained in a dark room indicates that high quality normal maps and surfaces are produced in these situations. The AAEs are small even with cast shadow in the input images.

**Outdoor Scene** Fig. 12 shows an outdoor example *HumanDoll* with the environment presented in Fig. 1(d). We put the equipment set on a trolley and move it around. The continuously changing lights cause difficulty for traditional methods but not for ours. We captured two sets of images with and without the presence of the sun. The latter was

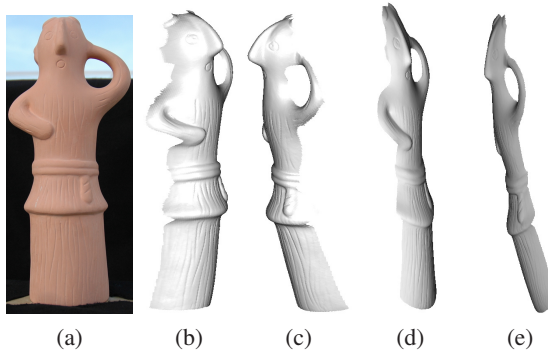


Figure 12. Outdoor example. (a) An input image. (b)-(c) Two views of the reconstructed surface using our method. (d)-(e) Two views of the surface result by the traditional method [26].

achieved under eaves. All these images are input to our system. Our result is with high quality.

**More Examples** Fig. 13 shows the results of *Kitten*, *Vase*, and *WineBottle*. In the *Kitten* example, the sun shines into the room through windows. This discontinuous extension of “ambient” is not allowed in other PS methods. The *Vase* is captured in a highly reflective room. High-quality surface normal and albedo are produced for all of the examples. Important structure as well as subtle details on the surfaces are properly preserved.

## 9. Conclusion

We have presented a method to mitigate data-collection restriction for photometric stereo. Rather than avoiding environment lighting, we treat both direct and indirect illumination as the same optical phenomenon, from which a simple and effective framework was derived. Our method uses a mirror sphere to calibrate environment lighting, making complex illumination no longer decisively harmful. Different environments were tested – from dark room to outdoor scenes. In many of them, decent surface can be produced.

We believe our framework can be naturally extended to other analytical shading models, such as the Ward’s model, which will be our future work.

## Acknowledgements

We thank Ping Tan for his comments on the method. This work is supported by a grant from the Research Grants Council of the Hong Kong SAR (project No. 413113).



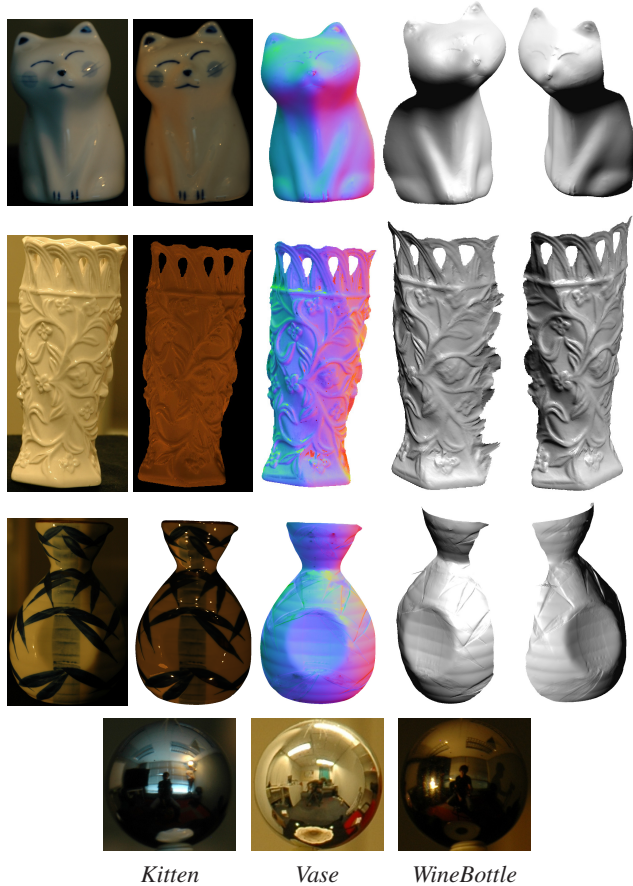


Figure 13. More examples. From top to bottom are *Kitten*, *Vase*, *WineBottle* and the corresponding mirror spheres. From left to right (for the top 3 examples) – input image, surface albedo, color coded normal map, surface results in two views.

## References

- [1] J. Ackermann, F. Langguth, S. Fuhrmann, and M. Goesele. Photometric stereo for outdoor webcams. In *CVPR*, pages 262–269, 2012.
- [2] A. Agrawal and R. Raskar. What is the range of surface reconstructions from a gradient field? In *ECCV*, 2006.
- [3] N. Alldrin, T. Zickler, and D. Kriegman. Photometric stereo with non-parametric and spatially-varying reflectance. In *CVPR*, 2008.
- [4] R. Basri, D. Jacobs, and I. Kemelmacher. Photometric stereo with general, unknown lighting. *Int. J. Comput. Vision*, 72(3):239–257, 2007.
- [5] H.-S. Chung and J. Jia. Efficient photometric stereo on glossy surfaces with wide specular lobes. *CVPR*, 2008.
- [6] E. Coleman-Jr. and R. Jain. Obtaining 3-dimensional shape of textured and specular surfaces using four-source photometry. *Comp. Graph. Image Proc.*, 18(4):309–328, 1982.
- [7] P. Debevec. Rendering synthetic objects into real scenes: bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *SIGGRAPH*, pages 189–198, 1998.
- [8] C. H. Esteban, G. Vogiatzis, and R. Cipolla. Multiview photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30:548–554, 2008.
- [9] D. B. Goldman, B. Curless, A. Hertzmann, and S. M. Seitz. Shape and spatially-varying brdfs from photometric stereo. In *ICCV05*, pages 341–348, 2005.
- [10] C. Hernandez, G. Vogiatzis, G. J. Brostow, B. Stenger, and R. Cipolla. Non-rigid photometric stereo with colored lights. *ICCV*, 2007.
- [11] A. Hertzmann and S. M. Seitz. Example-based photometric stereo: Shape reconstruction with general, varying brdfs. *IEEE Trans. PAMI*, 27(8):1254–1264, 2005.
- [12] T. Higo, Y. Matsushita, and K. Ikeuchi. Consensus photometric stereo. In *CVPR*, pages 1157 – 1164, 2010.
- [13] M. Holroyd, J. Lawrence, G. Humphreys, and T. Zickler. A photometric approach for estimating normals and tangents. *ACM Trans. Graph.*, 27(5), 2008.
- [14] J. Lim, J. Ho, M. Yang, and D. Kriegman. Passive photometric stereo from motion. In *ICCV05*, October 2005.
- [15] S. Nayar, K. Ikeuchi, and T. Kanade. Determining shape and reflectance of hybrid surfaces by photometric sampling. *IEEE Trans. Robo. Auto.*, 6(4):418–431, 1990.
- [16] D. Nehab, S. Rusinkiewicz, J. Davis, and R. Ramamoorthi. Efficiently combining positions and normals for precise 3D geometry. *ACM SIGGRAPH 2005*, 24(3), Aug. 2005.
- [17] F. E. Nicodemus. Directional reflectance and emissivity of an opaque surface. *Applied Optics*, 4(7):767–775, 1965.
- [18] J. Park, S. N. Sinha, Y. Matsushita, Y.-W. Tai, and I. S. Kweon. Multiview photometric stereo using planar mesh parameterization. In *ICCV*, pages 1161–1168, 2013.
- [19] J. Park, S. N. Sinha, Y. Matsushita, Y.-W. Tai, and I. S. Kweon. Calibrating a non-isotropic near point light source using a plane. In *CVPR*, 2014.
- [20] Y. Sato and K. Ikeuchi. Reflectance analysis under solar illumination. Technical report, Pittsburgh, USA, 1994.
- [21] L. Shen and P. Tan. Photometric stereo and weather estimation using internet images. *CVPR*, 2009.
- [22] B. Shi, Y. Matsushita, Y. Wei, C. Xu, and P. Tan. Self-calibrating photometric stereo. In *CVPR*, 2010.
- [23] F. Solomon and K. Ikeuchi. Extracting the shape and roughness of specular lobe objects using four light photometric stereo. *IEEE Trans. PAMI*, 18:449–454, 1996.
- [24] P. Tan, S. Lin, and L. Quan. Subpixel photometric stereo. *IEEE Trans. PAMI*, 30(8):1460–1471, 2008.
- [25] R. Woodham. Photometric method for determining surface orientation from multiple images. *Optical Eng.*, 19(1):139–144, 1980.
- [26] T.-P. Wu, K.-L. Tang, C.-K. Tang, and T.-T. Wong. Dense photometric stereo: A markov random field approach. *IEEE Trans. PAMI*, 28(11):1830–1846, 2006.
- [27] L.-F. Yu, S.-K. Yeung, Y.-W. Tai, D. Terzopoulos, and T. F. Chan. Outdoor photometric stereo. In *ICCP*, pages 1–8, 2013.
- [28] Z. Zhou, Z. Wu, and P. Tan. Multi-view photometric stereo with spatially varying isotropic materials. In *CVPR*, pages 1482–1489, 2013.