Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

A non-Lambertian photometric stereo under perspective projection*

Min LI†1, Chang-yu DIAO†‡2, Duan-qing XU1, Wei XING1, Dong-ming LU1
1College of Computer Science and Technology, Zhejiang University, Hangzhou 310027, China
2School of Art and Archeology, Zhejiang University, Hangzhou 310027, China
†E-mail: liminlim@126.com; dcy@zju.edu.cn

Received Mar. 19, 2019; Revision accepted Aug. 21, 2019; Crosschecked Nov. 12, 2019; Published online Dec. 21, 2019

Abstract: Under the perspective projection assumption, non-Lambertian photometric stereo is a highly non-linear problem. In this study, we present an optimized framework for reconstructing the surface normal and depth with non-Lambertian reflection models under perspective projection. By decomposing the images into diffuse and specular components, we compute the surface normal and reflectance simultaneously. We also propose a variational formulation that is robust and useful for surface reconstruction. The experiments show that our method accurately reconstructs both the surface shape and reflectance of colorful objects with non-Lambertian surfaces.

Key words: Photometric stereo; Three-dimensional reconstruction; Perspective projection; Image decomposition
https://doi.org/10.1631/FITEE.1900156 **CLC number:** TP391.4

1 Introduction

Photometric stereo is a technique for recovering the surface normal and reflectance using the reflectance model under different lighting conditions. Since Woodham (1979) first introduced photometric stereo for Lambertian scenes under orthogonal projection, the extension to real-world conditions, such as non-Lambertian appearances under perspective projection, has drawn significant interest. The reconstruction of the surface normal and depth can be used to provide highly detailed three-dimensional (3D) models in many applications, such as reverse engineering and digital preservation of cultural heritage.

The error of surface reflectance and depth reconstruction brought by the complex surface reflectance property is one of the most significant difficulties that prevent photometric stereo from being practical. Most previous photometric stereo algorithms focus only on computing surface normals with non-Lambertian scenes under orthogonal projection. One class of such methods assumes a perfect Lambertian model by adopting outlier rejection approaches to handle all non-Lambertian surfaces (Coleman and Jain, 1982; Solomon and Ikeuchi, 1996; Barsky and Petrou, 2003). With more input, recent methods (Wu et al., 2010; Ikehata et al., 2012) treat the problem as a matrix recovery task, augmenting matrix decomposition with an error matrix containing all outliers due to shadows, and specularity. This strategy is robust enough to handle shadows, but not applicable for specularity. A contrast class of methods (Goldman et al., 2010; Shi et al., 2014) attempts to adopt analytic models to fit a nonlinear analytic bi-directional reflectance distribution function (BRDF) to interpolate all the observed data, which may suffer from expensive computation resources

---

‡ Corresponding author

* Project supported by the Technological Program of Cultural Relics Preservation of Zhejiang Province, China, the Key Research and Development Program of Zhejiang Province, China (No. 2018C03051), and the National Standard Development Program of Cultural Relics Protection of China (No. 581250-T0170B)

ORCID: Min LI, http://orcid.org/0000-0003-4732-6457; Chang-yu DIAO, http://orcid.org/0000-0001-7744-0889

due to the nonlinear optimization of the inverse high dimensional reflectance models even under orthogonal projection. Another direction is to exploit physical properties such as reflectance isotropy and reciprocity. However, these methods are performed with sophisticated or specially designed setups, such as the light stage (Ghosh et al., 2009) or the ringlight (Zhou et al., 2013), which are not convenient to use outside the lab environment.

A promising way to overcome the above problems is to separate diffuse and specular components of the input images before computing the normal. Common ways to separate the images include employing color information (the dichromatic model (Shafer, 1985) and diffuse chromaticity (Yang et al., 2015)) and using polarization (Nayar et al., 1997; Wang et al., 2017). In this study, we present a new diffuse-specular separation strategy via a coordinated descent method based on the dichromatic model and simultaneously recover surface shapes and reflectance under perspective projection. Given the dichromatic model, we describe how to estimate both the surface normal and reflectance in an optimized framework based on color constancy. The framework allows us to eliminate specularity and leaves the ideally diffuse parts to compute the surface normal and reflectance. Instead of directly reconstructing the surface from the normal map (Agrawal et al., 2006b), we recover surface depth in a variational framework under perspective projection (Fig. 1), which is closer to real camera projection. In Section 6, we evaluate the surface normal and reflectance computed from the diffuse and specular components as well as the final mesh recovered using our reconstruction algorithm.

Our main contributions are summarized as follows: (1) We present a new approach for separating diffuse and specular components of input images using a coordinate descent method; (2) We simultaneously recover the geometry and photometric information; (3) We present an effective depth reconstruction algorithm under perspective projection.

2 Related work

2.1 Highlight detection and removal

Separating reflection components is one of the fundamental problems in computer vision and graphics, and is difficult due to the nature of the inherent physics. Many single- and multi-image methods have been developed to separate diffuse and specular reflections.

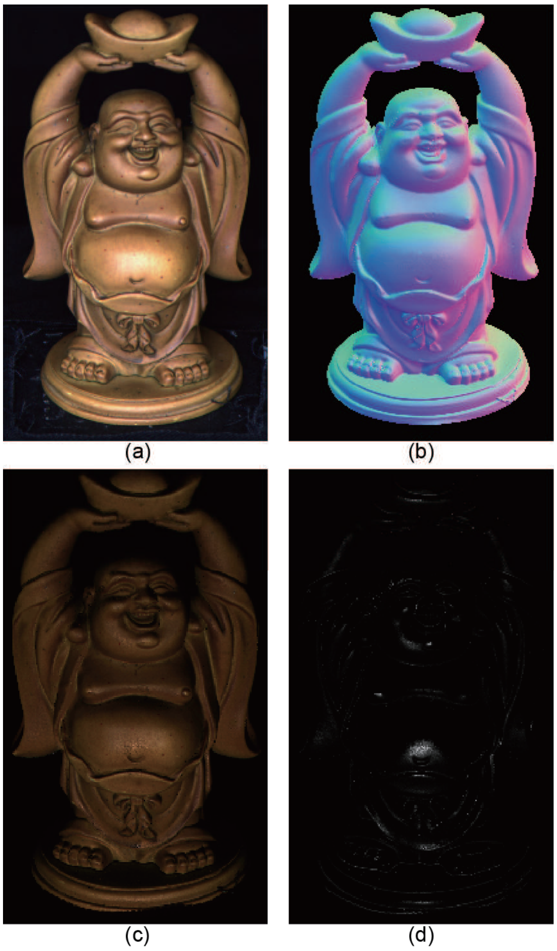


Fig. 1 Results of our method: (a) input images; (b) surface normal map; (c) diffuse image; (d) specular term

Most single-image methods employ additional information from natural images to eliminate the removal problem. Priors, such as long-tail distribution of gradients (Levin and Weiss, 2007), ghosting effect (Shih et al., 2015), and depth of field confidence maps (Wan et al., 2016), have been exploited. Recently, deep learning methods have been applied to this task (Wan et al., 2018; Yang et al., 2018; Zhang et al., 2018). Although traditional and deep learning techniques have made much progress for single-image reflection removal, there is still a significant margin for improvement due to the highly ill-posed nature of the problem.

Multiple images can provide more information to make the removal problem more traceable. Color spaces have been used to solve the highlight removal problem. Mallick et al. (2005) used a data-dependent rotation of the red-green-blue (RGB) color space to represent specular and diffuse components. Tan and Ikeuchi (2008) separated reflection components using a specular-to-diffuse mechanism given the diffuse maximum chromaticity. Yang et al. (2015) improved the method by developing a fast bilateral filtering approach to refine maximum chromaticity with neighboring pixels in real time. Recent works have focused on motion cues (Li and Brown, 2013; Guo et al., 2014; Sun et al., 2016; Han and Sim, 2018) by taking advantage of the difference in motion between the images of the same scene from different viewpoints.

Although the above multi-image methods work well for homogeneous and dichromatic surfaces in ideal cases, there are several significant limitations that make them difficult to use in practice. Most surfaces are textured, which violates the homogeneous assumption. A non-trivial pre-process like image segmentation is often introduced due to the homogeneous assumption, or the specular lobe must be narrow so that the specular and diffuse components of the dichromatic plane are distinct. Some of these limitations can be overcome using additional cues such as polarization (Nayar et al., 1997; Kong et al., 2012), varying focus (Schechner et al., 2000), and a flash-no-flash pair (Agrawal et al., 2005, 2006a). However, these methods still require narrow specular lobes or homogeneous surfaces.

2.2 Non-Lambertian photometric stereo

Photometric stereo is a technique of recovering both 3D shapes and surface materials from a series of images taken from a fixed viewpoint under multiple light directions. Photometric stereo methods typically rely significantly on knowledge of surface reflectance.

One class of these methods is to assume purely diffuse reflection models and consider specularity outliers because highlights of real object surfaces can be considered a linear combination of diffuse and specular lobes. Early researchers (Coleman and Jain, 1982; Solomon and Ikeuchi, 1996; Barsky and Petrou, 2003) proposed methods by selecting the three most Lambertian images out of four to estimate the surface normal. With enough input images, we can extract Lambertian surfaces by various of mathematical techniques such as graph cuts (Chandraker et al., 2007), random sample consensus (Sunkavalli et al., 2010), and expectation maximization (Wu et al., 2010). Recent statistical methods (Yu et al., 2010; Ikehata et al., 2012) assumed a low-rank observation matrix plus a sparse outliers matrix, inspired by methods of compressed sensing (Donoho, 2006). Ikehata et al. (2012) extended this method to a three-rank Lambertian constraint.

To effectively detect specularity, analytic reflectance models have been employed for non-Lambertian surfaces. Instead of removing specularity, analytical models are fitted to all the observed data, which leads to significant difference compared with statistical methods. Various reflectance models, such as the Blinn-Phong model (Esteban et al., 2008), the Torrance-Sparrow model (Georghiades, 2003), the Ward model (Chung and Jia, 2008), and the specular spike lobe (Yeung et al., 2015), have been incorporated to exploit all available data. By exploring more general reflectance properties, Ikehata and Aizawa (2014) employed a smooth and bivariate Bernstein polynomial function to model surface reflectance. Shi et al. (2014) proposed a bi-polynomial reflectance model to describe the low-frequency components of surface reflectance.

Besides the Lambertian model assumption, another typical simplified assumption is orthogonal projection. According to Tankus and Kiryati (2005), the perspective projection model has been shown to be a more realistic assumption. Until now there have been limited works dealing with the perspective projection model. Recently, Quéau et al. (2017, 2018) formulated non-convex variational models to directly recover the surface normal and depth with a little sacrifice of normal quality under perspective projection.

Approaches based on analytical physical models often introduce complicated non-linear optimizations, although they can be more useful for non-Lambertian surfaces. Approaches based on numerical methods usually assume a dominant Lambertian component, and are not applicable for most real surface materials. For the perspective projection assumption, the previous works often included a non-convex variational framework, which is time-consuming and non-linear. Recently, deep learning based methods (Santo et al., 2017; Ikehata, 2018) showed good performance on a public benchmark dataset. The significant constraints of these methods are that they need a lot of training data from various non-Lambertian surface reflectance and shapes, but they are effective only on certain objects. In this study, we present an approach that is similar to numerical approaches but without the assumption of a dominant component of surface materials. The proposed method can generate complete results without prior knowledge.

3 Dichromatic model

We know that reflectance at a surface point is described by BRDF, which defines how light is reflected from an opaque surface. Here, we consider it a five-dimensional function $f(\lambda, \boldsymbol{\theta})$ of wavelength and image formation, where $\boldsymbol{\theta} = (\theta_\mathrm{i}, \phi_\mathrm{i}, \theta_\mathrm{r}, \phi_\mathrm{r})$.

The real world is full of non-Lambertian materials. One of the widely used cases is the dichromatic model introduced by Shafer (1985). According to the dichromatic model, surface radiance is decomposed into two additive components, diffuse (body) reflectance and specular (interface) reflectance:

$$f(\lambda, \boldsymbol{\theta}) = f_\mathrm{d}(\boldsymbol{\theta}) m_\mathrm{d}(\lambda) + f_\mathrm{s}(\boldsymbol{\theta}). \tag{1}$$

Intuitively, the relative spectral power distribution m depends only on wavelength, and is independent of the surface geometry, whereas the geometric scale factors f_d and f_s are the opposite. Specifically, the geometric scale factor f_d for body reflection governs the proportion of diffuse light reflected from the object and depends only on the surface geometry.

To derive the image formation of a dichromatic surface, we again consider an observed surface point x with the surface normal $\boldsymbol{n}$ illuminated from light direction $\boldsymbol{l}$, and let $\boldsymbol{\theta} = (\theta_\mathrm{i}, \phi_\mathrm{i}, \theta_\mathrm{r}, \phi_\mathrm{r})$ represent illuminating and viewing directions under the local coordinate system. We assume that the sensitivity function of the camera sensor is linear and that the spectral power of the light source is represented as $E(\lambda)$. Under such assumptions, given BRDF at surface point x as in Eq. (1), the image formation is written as

$$I = m_\mathrm{d}(\lambda) f_\mathrm{d} E + E f_\mathrm{s}, \tag{2}$$

where $f_\mathrm{d} = \langle \boldsymbol{n}, \boldsymbol{l} \rangle$ is the shading term and m_d is the spectral reflectance of the observed surface point x as a diffuse part (Woodham, 1979).

4 Estimating surface normal and bidirectional reflectance distribution function parameters

We recover the surface normal and BRDF parameters using an optimized framework in this section. Our goal is to reduce the difference between the input images and the estimation results yielded by the correspondence BRDF model, and formulate the following energy function to minimize this difference:

$$F = \min_{f_\mathrm{d}, f_\mathrm{s}, m_\mathrm{d}, m_\mathrm{s}} \sum_{I_k \in \mathit{\Omega}} ||I_k - (m_\mathrm{d} f_\mathrm{d} E + f_\mathrm{s} E)||_2, \tag{3}$$

where $\mathit{\Omega}$ is defined as the set of image radiances in the same position under different light directions, $\mathit{\Omega} = \{I_1, I_2, \ldots, I_k\}$. $\mathit{\Omega}$ is related only to the body color. The function is minimized over the domain that contains a series of image intensities acquired under different light conditions at the same surface location. As a result, our method is not constrained by the uniform albedo assumption.

We now describe how to recover reflectance parameters and the surface normal. There are two steps in our optimized framework. First, we decompose the input image into a diffuse and a specular term. Then we adopt an iterative approach to find the normal $\boldsymbol{n}$ and m_d that yield the minimum result of Eq. (3).

4.1 Image decomposition based on the dichromatic model

The light distribution E is calibrated first, which means that our images are captured under the same light intensity ($[0, 0, 1]$). The significant characteristic of the dichromatic model is that the image radiance vectors of all points in the same surface lie in a two-dimensional subspace Q, spanned by the diffuse radiance vector and the light radiance vector, which have been widely used in prior work on color constancy (Finlayson and Schaefer, 2001a, 2001b). However, under our assumption, the image radiance vector of each point also lies in a two-dimensional subspace $Q \in \mathcal{R}^2$ spanned by the diffuse radiance vector $\boldsymbol{D}$ and the lighting intensity vector $\boldsymbol{E}$, respectively.

We adopt a solution similar to that used in Huynh and Robles-Kelly (2010), where we aim to decompose the images through different spectra with fixed lighting directions and different lighting intensities. To clarify, we describe how to obtain diffuse and specular terms in Algorithm 1 in detail.

Algorithm 1 Computing diffuse and specular terms

```
Require: Image $\boldsymbol{I}$ with image intensities $\{I_1, I_2, \ldots, I_k\}$ viewed from different lights with light intensities $\boldsymbol{E} = \{e_1, e_2, \ldots, e_k\}$ at pixel $\boldsymbol{x}$.
Ensure: $\boldsymbol{D}$, the diffuse reflectance of all points on the surface; $\boldsymbol{S}$, the specular term of all points on the surface; $\boldsymbol{f}_{\rm d}$, the shading term of all points on the surface.
for all $\boldsymbol{x} \in \boldsymbol{I}$ do
    $[\boldsymbol{D}, \boldsymbol{S}, \boldsymbol{f}_{\rm d}] \leftarrow \operatorname{argmin}_{\boldsymbol{D},\boldsymbol{S},\boldsymbol{f}_{\rm d}} F|_{\boldsymbol{E}}$
end for
return $\boldsymbol{D}, \boldsymbol{S}, \boldsymbol{f}_{\rm d}$
```

The subspace Q is obtained by decomposing $\boldsymbol{I}$ into two basis vectors $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ via singular value decomposition (SVD), where $Q = \operatorname{span}(\boldsymbol{z}_1, \boldsymbol{z}_2)$. Given $\boldsymbol{D}$ and $\boldsymbol{E}$ both lying in subspace Q, we parameterize $\boldsymbol{D} = v\boldsymbol{z}_1 + \boldsymbol{z}_2$ and $\boldsymbol{E} = w_1\boldsymbol{z}_1 + w_2\boldsymbol{z}_2$. Then we have

$$\begin{aligned}\boldsymbol{I} &= \boldsymbol{f}_{\rm d}\boldsymbol{D} + \boldsymbol{S}\boldsymbol{E} \\ &= (\boldsymbol{f}_{\rm d} \cdot v + \boldsymbol{S}w_1)\boldsymbol{z}_1 + (\boldsymbol{f}_{\rm d} + \boldsymbol{S}w_2)\boldsymbol{z}_2.\end{aligned} \tag{4}$$

For each point on the surface, a radiance vector $\boldsymbol{I} = (I_1, I_2, \ldots, I_k)$ is related to the shading term $\boldsymbol{f}_{\rm d} = (<\boldsymbol{n}, \boldsymbol{l}_1>, <\boldsymbol{n}, \boldsymbol{l}_2>, \cdots, <\boldsymbol{n}, \boldsymbol{l}_k>)$. It follows that $\boldsymbol{I}$ is mapped onto Q to obtain variables $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$:

$$\boldsymbol{I} = \boldsymbol{T}_1\boldsymbol{z}_1 + \boldsymbol{T}_2\boldsymbol{z}_2. \tag{5}$$

Based on Eqs. (4) and (5), we have

$$\boldsymbol{f}_{\rm d} = \frac{w_2\boldsymbol{T}_1 - w_1\boldsymbol{T}_2}{w_2 v - w_1}, \tag{6}$$

$$\boldsymbol{S} = \frac{\boldsymbol{T}_2 v - \boldsymbol{T}_2}{w_2 v - w_1}. \tag{7}$$

Because w_1 and w_2 are easily computed through $\boldsymbol{E}$, $\boldsymbol{z}_1$, and $\boldsymbol{z}_2$, the only remaining unknown variable is v. The optimization problem is converted into estimating v by solving a quadratic equation by substituting the above variables into Eq. (3) according to the algorithm in Huynh and Robles-Kelly (2010). Finally, the diffuse shading term $\boldsymbol{f}_{\rm d}$ is computed. It is easy to extract $m_{\rm d}$ and $m_{\rm s}$ from $\boldsymbol{D} = m_{\rm d}\boldsymbol{E}$ and $\boldsymbol{S} = m_{\rm s}\boldsymbol{E}$.

4.2 Normal and reflectance estimation

Given the lighting directions and intensities, the estimation of the surface normal via the straightforward matrix pseudo-inverse $[\cdot]^\dagger$ is described as

$$\boldsymbol{n} \propto \left[E_1\frac{\boldsymbol{l}_1}{\|\boldsymbol{l}_1\|}, E_2\frac{\boldsymbol{l}_2}{\|\boldsymbol{l}_2\|}, \ldots, E_k\frac{\boldsymbol{l}_k}{\|\boldsymbol{l}_k\|}\right]^\dagger [I_1, I_2, \ldots, I_k]. \tag{8}$$

Direct application of the algorithm previously described often generates poor results due to highlights that exist in the input images. The quality of the surface normal will be improved if we use the $\boldsymbol{f}_{\rm d}$ generated in Section 4.1. However, this will introduce outliers in regions containing achromatic points. Therefore, we propose an iterative approach to modify the initial normal. To identify the incorrect normal, we define specular confidence as $||(\boldsymbol{I} - \boldsymbol{I}')/\boldsymbol{I}||$, where $\boldsymbol{I}'$ includes the new image intensity $\boldsymbol{I}$ by incorporating parameters from each iteration, including normal $\boldsymbol{n}$, $\boldsymbol{D}$, and $\boldsymbol{S}$ in Eq. (2). Intuitively, the lower the confidence is, the more reliable the normal will be. We use the original radiance vector $\boldsymbol{I}$ to estimate the surface normal n_0 and compute the specular confidence for every point using n_0 as the initial input for the interactive procedures. At each iteration, we first compute the initial surface normal based on Eq. (8) using the radiance vector $\boldsymbol{f}_{\rm d}$ given by the previous iteration ($\boldsymbol{f}_{\rm d}$ used in the first iteration comes from the previous section) and compute the specular confidence for every point. For points whose specular confidence is decreasing, we arrange the intensity in each radiance vector $\boldsymbol{I}$ given by the previous iteration from high to low and discard the front 1% and the last 5% of the elements by pixel intensity. To obtain a new $\boldsymbol{f}_{\rm d}$, Algorithm 1 is employed using the updated radiance vector. The iterative process will stop until a specular confidence of more than 95% of points stops decreasing or there are no less than 60% of the pixels left in $\boldsymbol{I}$ due to the inherent property of photometric stereo. With this confidence strategy, we can estimate more reliable surface normals and specular parameters using only those confident pixels to eliminate the effect of highlight.

5 Depth reconstruction under perspective projection

In this section, we consider the depth reconstruction under perspective projection, which is closer to a real-world situation. We assume that $\hat{z}(x, y)$ denotes the depth function of object surfaces in a world Cartesian coordinate system whose origin is at the camera plane and the view direction $\boldsymbol{v}$ is $(0,0,1)$. If we project a 3D point X $(x,y,\hat{z})$ onto image point (u,v), then the depth is denoted as $z(u,v)$.

Given the surface normal, we recover a depth similar to Papadhimitri and Favaro (2013). Under perspective projection, the following relationship between image coordinate (u,v) and real-world coordinate $(x,y,\hat{z}(x,y))$ holds (Fig. 2):

$$\begin{cases} x=-u\cdot\hat{z}(x,y)/f, \\ y=-v\cdot\hat{z}(x,y)/f, \end{cases} \tag{9}$$

where f denotes the focal length, which is assumed to be known. So, the Lambertian reflectance equation can be expressed as a function of image coordinates only, which results in

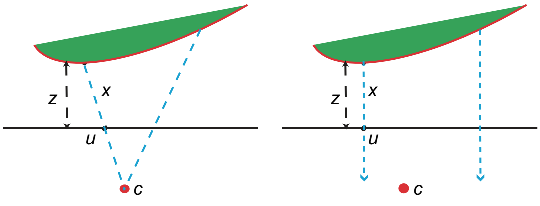
$$I(u,v)=\rho(u,v)\frac{(u-fp_{\rm s})\hat{p}+(v-fq_{\rm s})\hat{q}+1}{\|\boldsymbol{s}\|\sqrt{(u\hat{p}+v\hat{q}+1)^2+f^2(\hat{p}^2+\hat{q}^2)}}, \tag{10}$$

where $\hat{p}\triangleq z_u/z=\partial\ln z/\partial u$ and $\hat{q}\triangleq z_v/z=\partial\ln z/\partial v$ (Tankus and Kiryati, 2005; Papadhimitri and Favaro, 2013). Through the notation, the problem of recovering $z(u,v)$ from the image radiance equation reduces to the problem of recovering surface $\bar{z}=\ln z(u,v)$ from Eq. (10). Directly extracting normals under perspective projection is quite noisy (Tankus and Kiryati, 2005), so we transfer the normal $N=-(p,q,1)$ recovered in Section 4.2 to perspective projection format by $\hat{p}=-p/(z+up+vq)$ and $\hat{q}=-q/(z+up+vq)$.

The integration procedure of perspective photometric stereo is carried out using a Poisson solver (Agrawal et al., 2006b), which adds an additional process by taking the exponent of the integration results. However, this often fails due to the non-integrity of the gradient field estimated from the photometric stereo in the presence of outliers and noises. So, to integrate surfaces from the 2D gradient field, the variational framework is adopted as

$$\begin{aligned} E_{\rm TV}(\bar{z})=\iint\limits_{(u,v)\in\varOmega}&\Big(||\nabla\bar{z}(u,v)-\hat{\boldsymbol{g}}(u,v)|| \\ &+\lambda(u,v)\left(\bar{z}(u,v)-\hat{z}^0(u,v)\right)^2\Big){\rm d}u{\rm d}v, \end{aligned} \tag{11}$$

where $\bar{z}^0$ is the logarithm of the initial rough depth and $\hat{\boldsymbol{g}}(u,v)=<\hat{p},\hat{q}>$. In this study, we find that shadows and noises generate more noise than outliers, so we pick the l_2 norm, which shows robustness when facing the noise of the gradient field $\hat{\boldsymbol{g}}$ generated from the natural logarithm of depth z. The weight λ is to control the smoothness of the surface. Eq. (11) can be solved by global optimization, such as an alternating direction method of multipliers (Gabay and Mercier, 1976), considering that it is convex over the whole surface.



(a) (b)

Fig. 2 Different projections: (a) perspective model; (b) orthogonal projection model

6 Experiments

To evaluate our method, we tested it on real images captured by setups with different light configurations, including the data of DiLiGenT (Shi et al., 2019) and our own data (Fig. 3 and Table 1). All

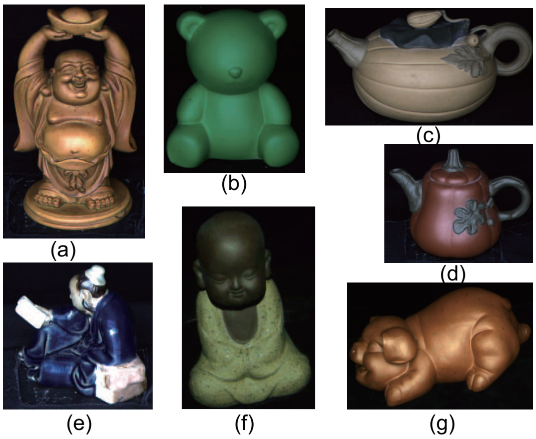

(a) (b) (c) (d) (e) (f) (g)

Fig. 3 Data used in the experiments: (a) BUDDHA; (b) BEAR; (c) POT1; (d) POT2; (e) READING; (f) SBUDDHA; (g) PIG

Table 1 Data used in the experiments

Dataset	Data	Number of LEDs	Image resolution
DiLiGenT	BUDDHA BEAR POT1 POT2 READING	96	512 × 612
Our data	SBUDDHA PIG	72	480 × 640

experiments were performed on a work-station with Intel Core2 i7-4790 (3.60 GHz) and 16 GB RAM, and implemented in MATLAB.

As for ground truth used in the comparison, the DiLiGent benchmark presents the "normal ground truth" for all data. The details of capturing PIG and SBUDDHA were almost the same as those in Zhou et al. (2013) except that we used a cheaper commercial camera with image resolution 480 × 640 and two circles of ring light with radii of 75 mm and 150 mm. Furthermore, we used a commercial 3D scanner to acquire the ground truth shape for PIG. The relative calibration information, including intrinsic and extrinsic camera parameters, lighting directions, and intensities, was provided affiliated with the dataset.

Through all experiments, our method was compared mainly against five representative photometric stereo methods, including the baseline (LS) method (Woodham, 1979), IA14 (Ikehata and Aizawa, 2014), ST14 (Shi et al., 2014) that focused only on normal estimation, and YW17 (Quéau et al., 2017) and YW18 (Quéau et al., 2018) where both surface normal and depth are compared. In Section 6.1, we also evaluated two other most recent methods, HS15 (Han and Shen, 2015) and SH17 (Shen et al., 2017), for normal evaluation. For all competitive methods, we used the parameters provided in the original codes or suggested by the original papers. We quantitatively evaluated the two key steps of our method: surface normal computation and depth reconstruction. We also showed our intermediate results and evaluated varying lighting distributions to prove the robustness of our algorithm.

Our normal evaluation criteria were based on the statistics of the angular error. For each pixel, the angular error was calculated as $\arccos(\boldsymbol{n}^0, \boldsymbol{n})$ in degrees, where $\boldsymbol{n}^0$ and $\boldsymbol{n}$ are the "ground truth" and estimated normal, respectively. We adopted the mean angular error (MAE) as the metric, which is for normal estimation. The shape error was calculated as $\|\boldsymbol{X}^0 - \boldsymbol{X}\|$ in millimeters, where $\boldsymbol{X}^0$ and $\boldsymbol{X}$ are the "ground truth" and estimated 3D position, respectively. We also employed the mean shape error (MSE) as the mesh estimation metric.

6.1 Evaluation of normal computation

We first evaluated the performance of the normal estimation method as described in Section 4. We used all data for the overall performance. For fair comparison, we directly used normal maps given by the benchmark for IA14 and ST14 of BUDDHA, BEAR, POT1, POT2, and READING. Results of SBUDDHA and PIG were computed from source codes given by Shi et al. (2019).

6.1.1 A position threshold

It is argued that shadows and specular highlights are the main problems that cause estimation errors for photometric stereo. Because shadows and highlights are often associated with dark and bright pixels, we could simply sort pixels by their intensities and discard shadows and highlights by a simple threshold separately. Such a strategy has been used in many non-Lambertian methods (ST14 and IA14) as initialization to achieve the best performance. In our experiments, we used an iterative strategy to discard shadows and highlights according to Section 4.2. When the specular confidence of all points stopped decreasing, the iteration stopped. Fig. 4 and Table 2 show the first five iterations of the normal estimation. In each iteration, we found that the best normal was achieved when $T_\mathrm{f} = 1\%$ and $T_\mathrm{l} = 5\%$. This was consistent with our assumption that we used highlight information for surface reflectance reconstruction and it was reasonable to save more bright pixels for normal estimation. We found that the iteration stopped at the third or fourth iteration. In the following, we take the results of the fourth iteration as the final results.

For the computation of specular confidence, we selected only reliable image intensities by discarding the dark and bright pixels. We set the position thresholds for shadow and highlight among $\boldsymbol{T}_\mathrm{low} = [0\%, 25\%, 45\%]$ and $\boldsymbol{T}_\mathrm{high} = [100\%, 75\%, 55\%]$. Table 3 shows the normal angular errors using different thresholds, indicating that $[25\%, 75\%]$ produces comparable results to the state-of-the-art methods.

Table 2 MAE of normals for the intermediate results in Fig. 4 with different numbers of iterations

Data	MAE				
	1	2	3	4	5
BUDDHA	10.32	9.41	**8.74**	8.74	8.74
BEAR	6.51	5.90	5.24	**5.10**	5.10
POT1	6.74	6.23	**6.04**	6.04	6.04
POT2	8.11	7.60	**7.27**	7.27	7.27
READING	14.82	12.31	12.06	**11.99**	11.99

The bold fonts represent the MAE of final normals

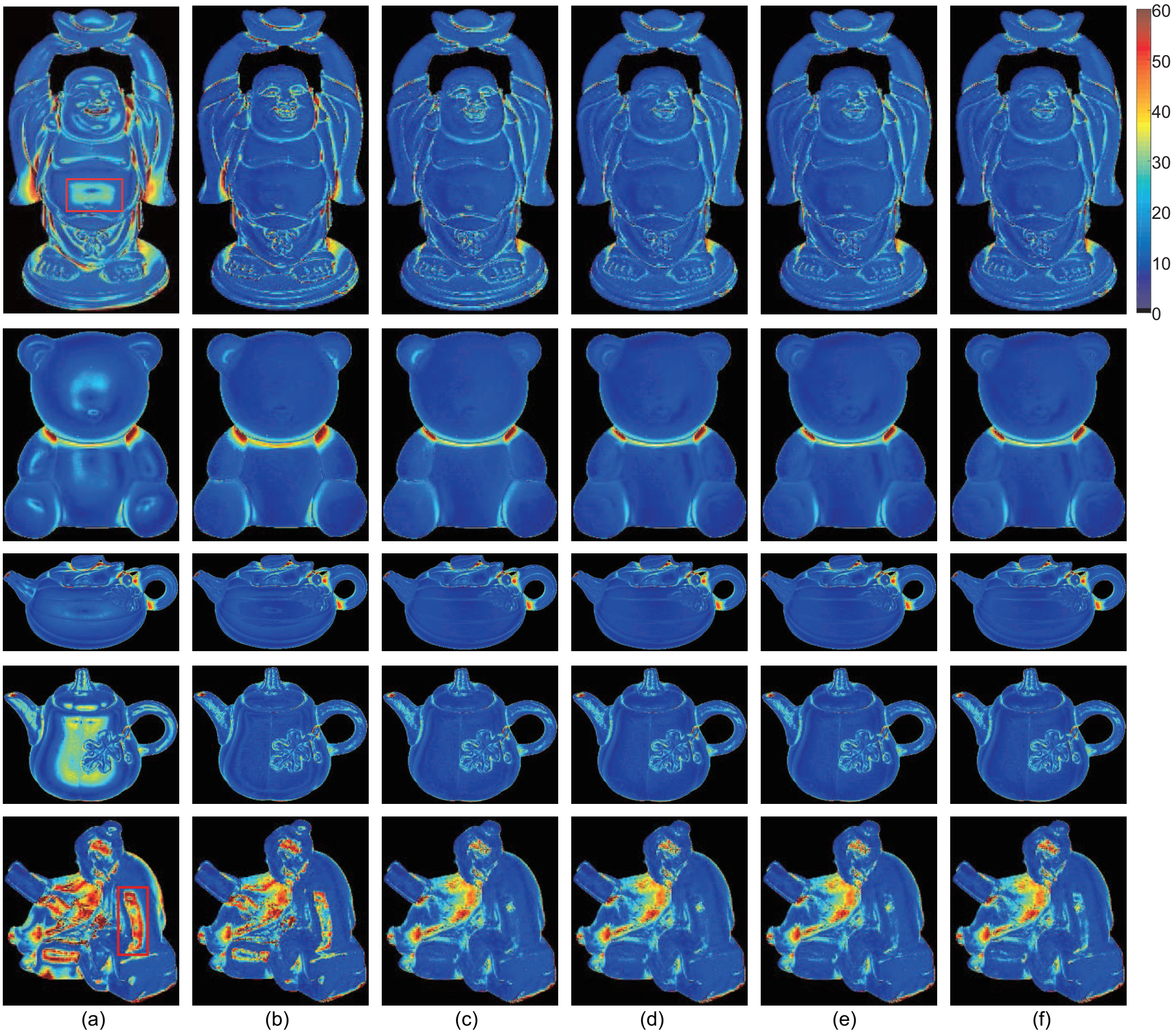


Fig. 4 Intermediate mean normal angular error map: (a) MAE maps of normals for LS; (b) normal error maps of the 1$^{\text{st}}$ iteration; (c) normal error maps of the 2$^{\text{nd}}$ iteration; (d) normal error maps of the 3$^{\text{rd}}$ iteration; (e) normal error maps of the 4$^{\text{th}}$ iteration; (f) normal error maps of the 5$^{\text{th}}$ iteration. References to color refer to the online version of this figure

Table 3 MAE of normals using different thresholds for computing specular confidence

Data	MAE		
	[0%, 100%]	[25%, 75%]	[45%, 55%]
BUDDHA	8.99	**8.74**	9.01
BEAR	5.65	**5.10**	5.37
POT1	6.08	**6.04**	6.09
POT2	7.34	**7.27**	7.47
PIG	13.58	**11.99**	12.52

The bold fonts represent the best normal results

We chose [25%, 75%] through the whole experiments. A proper threshold is effective for objects with strong specularity, for example, the red box areas of BUDDHA and READING in Fig. 4.

6.1.2 Number of input images

It is claimed in previous papers that several photometric stereo methods, IA14 and ST14, need hundreds of input images to perform analytical modeling or statistical analysis, where low-frequency analysis is adopted by discarding high intensity pixels. Conversely, our method uses highlights by separating them from the diffuse information rather than discarding highlight points directly. Hence, our method is expected to work well on fewer input images compared with those used in IA14 and ST14. For fair comparison, we also evaluated the performance of LS, YW17, and YW18. The experiments were conducted with input images of 6, 12, 18, 24, 30, 36, 42, 48, and 72, separately, on BUDDHA, BEAR, POT1, POT2, and READING. The light directions were set randomly. To eliminate the uncertainty of the random input, we report the average results under five different light conditions of each number of input images. The results are illustrated in Fig. 5. The curves indicate that our method always reaches the final result with around 40 input images, whereas other methods are still at a relatively high level. It can be found that YW17 and YW18 achieve slightly better performance than others when the number of input images is smaller than 30. This is reasonable because both YW17 and YW18 depend on the Lambertian reflectance model assumption but cannot employ adequate information when the input number is growing. Other robust methods, such as ST14 and IA14, need more input images than ours to achieve the best results.

6.1.3 Evaluation on the benchmark

To validate the robustness of our algorithm, we selected typical types of objects, POT1 with a rough surface, PIG and READING with some specular spikes, BEAR and BUDDHA with broad and soft specular lobes, and POT2 with spatially varying materials. Fig. 6 shows the normal maps and corresponding error maps for our method, handling the specular highlights identically by the existing robust photometric stereo methods rather than discarding them as outliers. Results showed that our method generates large errors on concave areas. Like other pixel-wise methods, we did not make assumptions on cast shadows or inter-reflection. Then we compared the proposed method with the state-of-the-art methods. Table 4 shows the MAE of our method and state-of-the-art methods. For each object, the best result is highlighted with bold font. As shown in Table 4, our method achieves the best results on all objects. The comparison shows that our method achieves high accuracy for objects consisting of various materials with different specular effects.

6.1.4 Evaluation on our data

To further evaluate our method, we tested data captured by ourselves, including SBUDDHA and PIG. We used the same parameter settings as for the benchmark data. Table 4 shows that our method achieves the best result on data compared with state-of-the-art methods. Fig. 7 further illustrates that our method performs best on the highlighted areas of SBUDDHA and PIG. The observation is consistent with our conclusion that we use highlights rather than discarding them as outliers to estimate the surface normal. The results also prove that our method is robust enough to estimate normals from images captured by cheap and easily acquirable commercial cameras.

Besides the surface normal, we evaluated the accuracy of our computed reflectance. We adapted the usual way of rendering images in photometric stereo. Figs. 8a and 8b look really similar, which

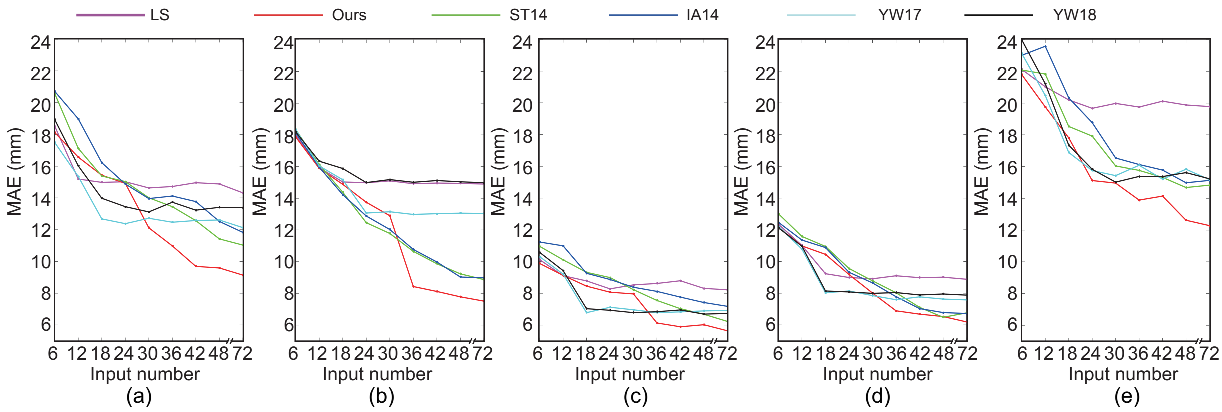


Fig. 5 Normal error distributions for the related methods under different illumination conditions: (a) BUDDHA; (b) BEAR; (c) POT1; (d) POT2; (e) READING

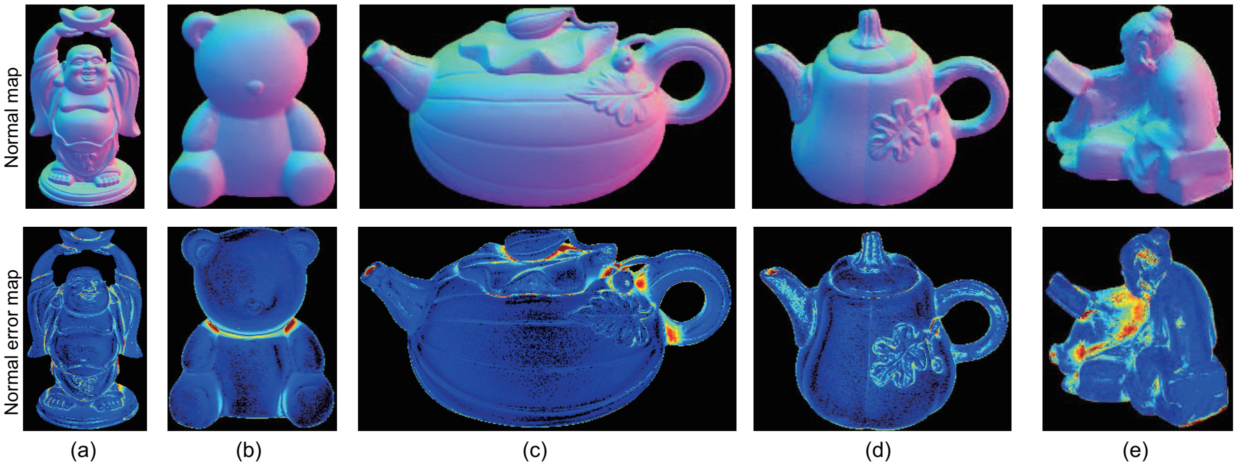


Fig. 6 Final results of normal evaluation: (a) BUDDHA; (b) BEAR; (c) POT1; (d) POT2; (e) READING

Table 4 MAE of normals of objects used in our experiments

Data	MAE							
	LS	Ours	IA14	ST14	HS15	SH17	YW17	YW18
BUDDHA	14.92	**8.74**	10.47	10.60	12.29	9.30	12.01	13.43
BEAR	8.39	**5.10**	7.11	6.12	5.12	5.31	6.89	6.71
POT1	8.89	**6.04**	6.64	6.51	6.89	7.28	7.61	7.87
POT2	14.65	**7.27**	8.77	8.78	9.80	8.43	13.04	13.07
READING	19.80	**11.99**	14.19	13.63	14.56	13.00	15.74	15.04
SBUDDHA	17.22	**12.02**	18.74	13.64	14.28	14.34	15.06	17.04
PIG	23.30	**14.41**	22.51	17.53	19.34	17.82	19.40	16.54

The bold fonts represent the best results of objects

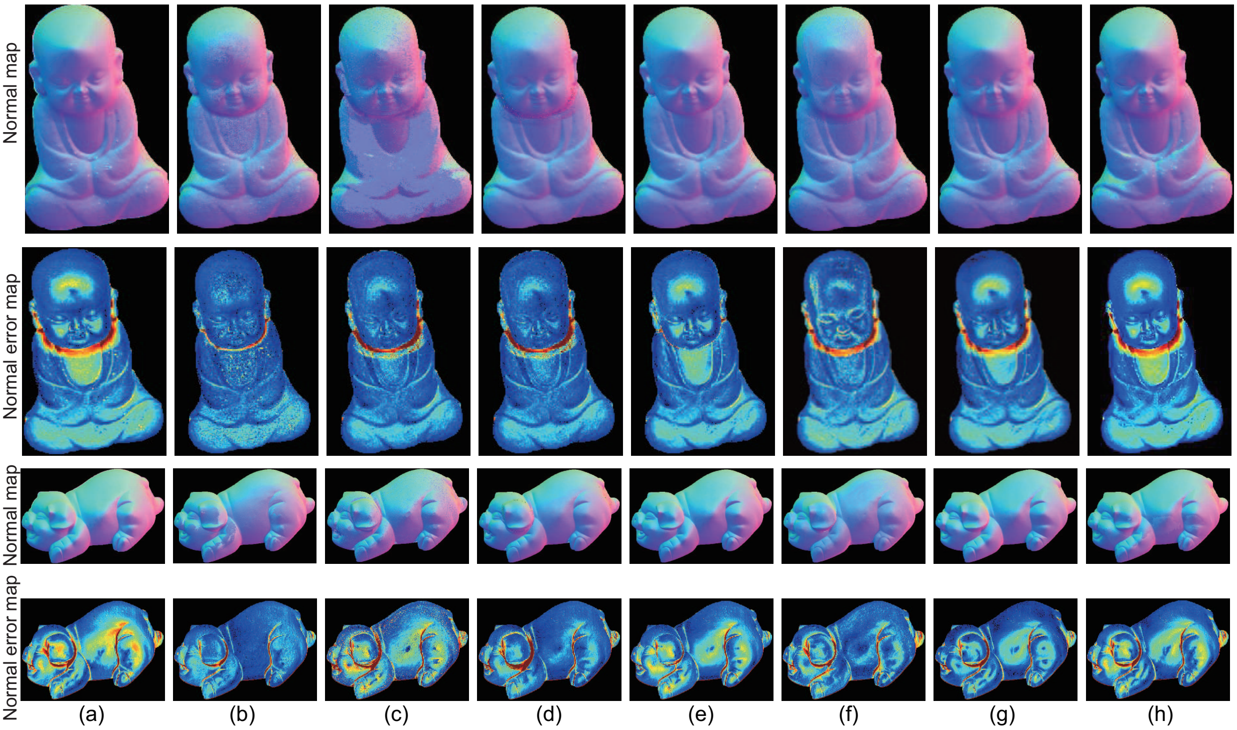


Fig. 7 Normal maps and normal error maps of SBUDDHA and PIG for LS (a), ours (b), IA14 (c), ST14 (d), HS15 (e), SH17 (f), YW17 (g), and YW18 (h)

Fig. 8 Images rendered using our computed BRDF parameters: (a) input images; (b) rendered images

means that our strategy of computing a reflectance map in Section 4 is very effective.

6.2 Evaluation of mesh estimation

In this subsection, we present the BUDDHA, BEAR, POT1, POT2, and PIG shapes, covering a wide range of shapes from smoothly curved surfaces to complicated geometry. Among them, PIG is captured under a field of view (FOV) of about 30°, indicating obvious perspective effects. To evaluate the shape errors accurately, we computed the depth map with a rough initial surface depth 1000 mm, using the depth reconstruction algorithms presented in Section 5. It was observed that the proposed depth reconstruction method was less sensitive to parameter λ in a wide range $[10^{-6}, 10^{-3}]$. In the following, without extra statement, we chose $\lambda = 10^{-4}$.

We compared the quality of the final shapes with those obtained using the state-of-the-art methods presented by Quéau et al. (2017, 2018), both of which focused on tackling the photometric stereo problem in the presence of near-light perspective effects, similar to ours.

Fig. 9 shows our results for the surface shapes. Our method reconstructs most surface details, especially in the area with specular reflection. Considering BUDDHA and PIG with the most complicated geometries, the meshes were quite well aligned with the ground truth. One reason is that the results of normals computed in Section 6.1 generate a more reliable integration path. Another reason is that the process of depth reconstruction is more robust because the L_1 regularization term is effective in discarding those outliers caused by the non-integrable specular area. In our experiments, most of the large shape errors appeared at the concave carving with strong inter-reflection. This is because the noisy normal cannot generate the right integrating path, causing significant errors in areas where the gradient field tends to infinity. The color coded shape error shows the phenomenon clearly.

Fig. 10 illustrates the results of YW18 and YW17 on BUDDHA and PIG because they contain the most complicated geometry. The mean shape errors of the corresponding methods were 3.57 mm and 3.53 mm for BUDDHA and 3.76 mm and 3.84 mm for PIG. Both methods performed poorly when reconstructing surfaces in the highlight area (red box areas in Fig. 10). This is not surprising because no method can detect highlight effectively, which introduces more noise for the gradient field and generates an incorrect integration path for depth reconstruction. Fig. 11 shows the detailed meshes of YW18, YW17, and our method. Our results are closest to the ground truth, further proving that our method performs better when dealing with highlighted areas.

To test the robustness of our algorithm in detail, we evaluated all three methods using both orthogonal and perspective methods. We call our depth reconstruction algorithm a perspective algorithm and the Poisson depth reconstruction (Agrawal et al., 2006b) an orthogonal algorithm. We used both perspective and orthogonal algorithms for BUDDHA,

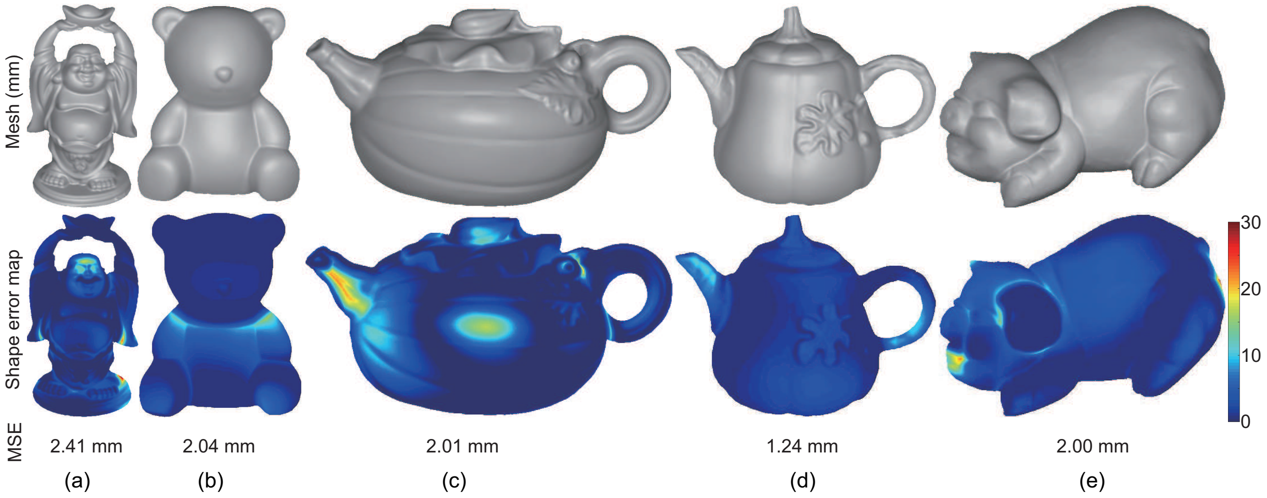


Fig. 9 MSE of BUDDHA (a), BEAR (b), POT1 (c), POT2 (d), and PIG (e)

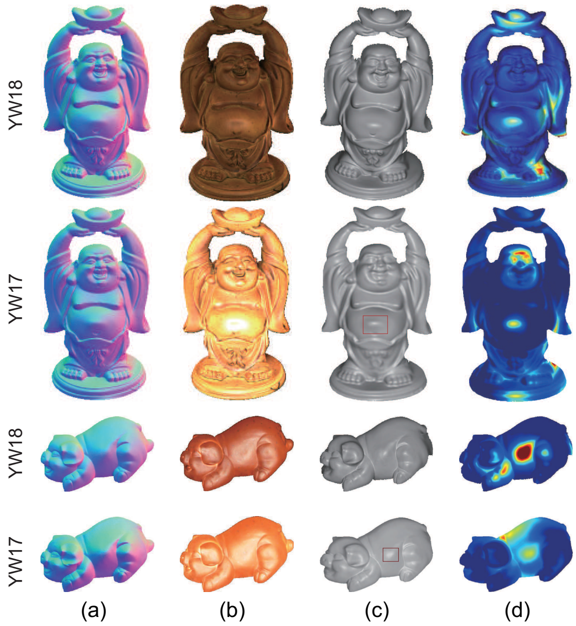


Fig. 10 Results of the methods of YW18 and YW17: (a) normal; (b) rendered image; (c) mesh; (d) shape error map. References to color refer to the online version of this figure

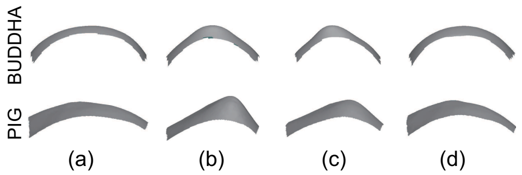


Fig. 11 Detailed meshes of the red box areas in Fig. 10: (a) ground truth; (b) YW18; (c) YW17; (d) Ours

BEAR, POT1, POT2, and PIG to evaluate final shapes. We compared our method with YW17 and YW18, which were also used to reconstruct the surface under perspective projection. Normal maps estimated from the three methods were considered input for orthogonal projection. Results of MSE are presented in Fig. 12. We may have the following conclusions: (1) The perspective algorithm improves the quality of the final shapes compared with the orthogonal shapes; (2) When perspective effectiveness is more obvious (PIG), our method performs better. We present Table 5 for a detailed explanation,

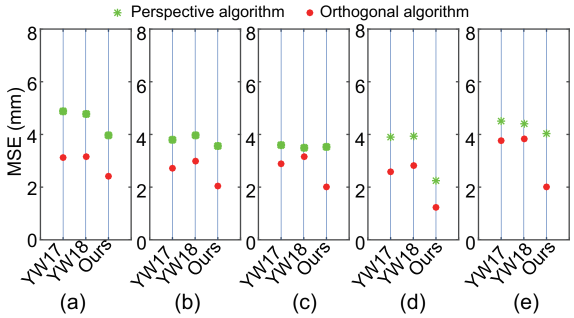


Fig. 12 MSE of the objects: (a) BUDDHA; (b) BEAR; (c) POT1; (d) POT2; (e) PIG

Table 5 Mesh ratio of different objects with different methods

Data	Mesh ratio		
	YW17	YW18	Ours
BUDDHA	1.35	1.33	1.43
BEAR	1.40	1.33	1.74
POT1	1.23	1.10	1.75
POT2	1.42	1.34	1.78
PIG	1.20	1.15	**2.01**

The bold font represents the best result where we define the orthogonal mesh divided by the perspective mesh as a mesh ratio. Large mesh ratios indicate that our method provides better results.

7 Conclusions and future work

We present a new photometric stereo method to capture both surface normal and reflectance of real-world objects. Our method is general and works with non-Lambertian BRDFs. The errors of captured surface normal and meshes are comparable to those of the state-of-the-art methods. The results prove that our method is effective in extracting specularity from object surfaces, which is quite suitable for a number of applications such as digital preservation of cultural heritage. For example, it is inspiring to apply our method to exhibit museum collections with both accurate shapes and vivid surfaces online.

Our method has a few limitations, however. First, it cannot model anisotropic materials, transparent objects, or mirror surfaces. Second, without initial 3D points, we are not able to reconstruct accurate 3D models. An interesting direction is to combine other 3D reconstruction methods, such as multi-view stereo (Galliani et al., 2015), with our work to simultaneously acquire the perfect 3D shape and the reflectance. It is also planned to incorporate other computer vision algorithms, such as image classification methods (Peng et al., 2018; He et al., 2019a, 2019b; He and Peng, 2019), with ours to analyze the local regions of objects and parts.

Compliance with ethics guidelines

Min LI, Chang-yu DIAO, Duan-qing XU, Wei XING, and Dong-ming LU declare that they have no conflict of interest.

References

Agrawal A, Raskar R, Nayar SK, et al., 2005. Removing photography artifacts using gradient projection and flash-exposure sampling. *ACM Trans Graph*, 24(3):828-835. https://doi.org/10.1145/1073204.1073269

Agrawal A, Raskar R, Chellappa R, 2006a. Edge suppression by gradient field transformation using cross-projection tensors. IEEE Computer Society Conf on Computer Vision and Pattern Recognition, p.2301-2308. https://doi.org/10.1109/CVPR.2006.106

Agrawal A, Raskar R, Chellappa R, 2006b. What is the range of surface reconstructions from a gradient field? 9th European Conf on Computer Vision, p.578-591. https://doi.org/10.1007/11744023_45

Barsky S, Petrou M, 2003. The 4-source photometric stereo technique for three-dimensional surfaces in the presence of highlights and shadows. *IEEE Trans Patt Anal Mach Intell*, 25(10):1239-1252. https://doi.org/10.1109/TPAMI.2003.1233898

Chandraker M, Agarwal S, Kriegman D, 2007. Shadowcuts: photometric stereo with shadows. IEEE Conf on Computer Vision and Pattern Recognition, p.1-8. https://doi.org/10.1109/CVPR.2007.383288

Chung HS, Jia JY, 2008. Efficient photometric stereo on glossy surfaces with wide specular lobes. IEEE Conf on Computer Vision and Pattern Recognition, p.1-8. https://doi.org/10.1109/CVPR.2008.4587771

Coleman ENJr, Jain R, 1982. Obtaining 3-dimensional shape of textured and specular surfaces using four-source photometry. *Comput Graph Image Process*, 18(4):309-328. https://doi.org/10.1016/0146-664X(82)90001-6

Donoho DL, 2006. Compressed sensing. *IEEE Trans Inform Theory*, 52(4):1289-1306. https://doi.org/10.1109/TIT.2006.871582

Esteban CH, Vogiatzis G, Cipolla R, 2008. Multiview photometric stereo. *IEEE Trans Patt Anal Mach Intell*, 30(3):548-554. https://doi.org/10.1109/TPAMI.2007.70820

Finlayson GD, Schaefer G, 2001a. Convex and non-convex illuminant constraints for dichromatic colour constancy. Proc IEEE Computer Society Conf on Computer Vision and Pattern Recognition, p.598-604 https://doi.org/10.1109/CVPR.2001.990528

Finlayson GD, Schaefer G, 2001b. Solving for colour constancy using a constrained dichromatic reflection model. *Int J Comput Vis*, 42(3):127-144. https://doi.org/10.1023/A:1011120214885

Gabay D, Mercier B, 1976. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Comput Math Appl*, 2(1):17-40. https://doi.org/10.1016/0898-1221(76)90003-1

Galliani S, Lasinger K, Schindler K, 2015. Massively parallel multiview stereopsis by surface normal diffusion. Proc IEEE Int Conf on Computer Vision, p.873-881. https://doi.org/10.1109/ICCV.2015.106

Georghiades AS, 2003. Incorporating the Torrance and Sparrow model of reflectance in uncalibrated photometric stereo. Proc 9th IEEE Int Conf on Computer Vision, Article 816. https://doi.org/10.1109/ICCV.2003.1238432

Ghosh A, Chen TB, Peers P, et al., 2009. Estimating specular roughness and anisotropy from second order spherical gradient illumination. *Comput Graph For*, 28(4):1161-1170. https://doi.org/10.1111/j.1467-8659.2009.01493.x

Goldman DB, Curless B, Hertzmann A, et al., 2010. Shape and spatially-varying BRDFs from photometric stereo. *IEEE Trans Patt Anal Mach Intell*, 32(6):1060-1071. https://doi.org/10.1109/TPAMI.2009.102

Guo XJ, Cao XC, Ma Y, 2014. Robust separation of reflection from multiple images. Proc IEEE Conf on Computer Vision and Pattern Recognition, p.2187-2194. https://doi.org/10.1109/CVPR.2014.281

Han BJ, Sim JY, 2018. Glass reflection removal using co-saliency-based image alignment and low-rank matrix completion in gradient domain. *IEEE Trans Image Process*, 27(10):4873-4888. https://doi.org/10.1109/TIP.2018.2849880

Han TQ, Shen HL, 2015. Photometric stereo for general BRDFs via reflection sparsity modeling. *IEEE Trans Image Process*, 24(12):4888-4903. https://doi.org/10.1109/TIP.2015.2471081

He XT, Peng YX, 2019. Fine-grained visual-textual representation learning. *IEEE Trans Circ Syst Video Technol*, online. https://doi.org/10.1109/TCSVT.2019.2892802

He XT, Peng YX, Zhao JJ, 2019a. Fast fine-grained image classification via weakly supervised discriminative localization. *IEEE Trans Circ Syst Video Technol*, 29(5):1394-1407. https://doi.org/10.1109/TCSVT.2018.2834480

He XT, Peng YX, Zhao JJ, 2019b. Which and how many regions to gaze: focus discriminative regions for fine-grained visual categorization. *Int J Comput Vis*, 127(9):1235-1255. https://doi.org/10.1007/s11263-019-01176-2

Huynh CP, Robles-Kelly A, 2010. A solution of the dichromatic model for multispectral photometric invariance. *Int J Comput Vis*, 90(1):1-27. https://doi.org/10.1007/s11263-010-0333-y

Ikehata S, 2018. CNN-PS: CNN-based photometric stereo for general non-convex surfaces. European Conf on Computer Vision, p.3-19. https://doi.org/10.1007/978-3-030-01267-0_1

Ikehata S, Aizawa K, 2014. Photometric stereo using constrained bivariate regression for general isotropic surfaces. Proc IEEE Conf on Computer Vision and Pattern Recognition, p.2179-2186. https://doi.org/10.1109/CVPR.2014.280

Ikehata S, Wipf D, Matsushita Y, et al., 2012. Robust photometric stereo using sparse regression. IEEE Conf on Computer Vision and Pattern Recognition, p.318-325. https://doi.org/10.1109/CVPR.2012.6247691

Kong N, Tai YW, Shin SY, 2012. A physically-based approach to reflection separation. IEEE Conf on Computer Vision and Pattern Recognition, p.9-16. https://doi.org/10.1109/CVPR.2012.6247652

Levin A, Weiss Y, 2007. User assisted separation of reflections from a single image using a sparsity prior. *IEEE Trans Patt Anal Mach Intell*, 29(9):1647-1654. https://doi.org/10.1109/TPAMI.2007.1106

Li Y, Brown MS, 2013. Exploiting reflection change for automatic reflection removal. Proc IEEE Int Conf on Computer Vision, p.2432-2439. https://doi.org/10.1109/ICCV.2013.302

Mallick SP, Zickler TE, Kriegman DJ, et al., 2005. Beyond Lambert: reconstructing specular surfaces using color. IEEE Computer Society Conf on Computer Vision and Pattern Recognition, p.619-626. https://doi.org/10.1109/CVPR.2005.88

Nayar SK, Fang XS, Boult T, 1997. Separation of reflection components using color and polarization. *Int J Comput Vis*, 21(3):163-186. https://doi.org/10.1023/A:1007937815113

Papadhimitri T, Favaro P, 2013. A new perspective on uncalibrated photometric stereo. Proc IEEE Conf on Computer Vision and Pattern Recognition, p.1474-1481. https://doi.org/10.1109/CVPR.2013.194

Peng YX, He XT, Zhao JJ, 2018. Object-part attention model for fine-grained image classification. *IEEE Trans Image Process*, 27(3):1487-1500. https://doi.org/10.1109/TIP.2017.2774041

Quéau Y, Wu T, Lauze F, et al., 2017. A non-convex variational approach to photometric stereo under inaccurate lighting. Proc IEEE Conf on Computer Vision and Pattern Recognition, p.350-359. https://doi.org/10.1109/CVPR.2017.45

Quéau Y, Durix B, Wu T, et al., 2018. LED-based photometric stereo: modeling, calibration and numerical solution. *J Math Imag Vis*, 60(3):313-340. https://doi.org/10.1007/s10851-017-0761-1

Santo H, Samejima M, Sugano Y, et al., 2017. Deep photometric stereo network. Proc IEEE Int Conf on Computer Vision, p.501-509. https://doi.org/10.1109/ICCVW.2017.66

Schechner YY, Kiryati N, Basri R, 2000. Separation of transparent layers using focus. *Int J Comput Vis*, 39(1):25-39. https://doi.org/10.1023/A:1008166017466

Shafer SA, 1985. Using color to separate reflection components. *Color Res Appl*, 10(4):210-218. https://doi.org/10.1002/col.5080100409

Shen HL, Han TQ, Li CG, 2017. Efficient photometric stereo using kernel regression. *IEEE Trans Image Process*, 26(1):439-451. https://doi.org/10.1109/TIP.2016.2627805

Shi BX, Tan P, Matsushita Y, et al., 2014. Bi-polynomial modeling of low-frequency reflectances. *IEEE Trans Patt Anal Mach Intell*, 36(6):1078-1091. https://doi.org/10.1109/TPAMI.2013.196

Shi BX, Mo ZP, Wu Z, et al., 2019. A benchmark dataset and evaluation for non-Lambertian and uncalibrated photometric stereo. *IEEE Trans Patt Anal Mach Intell*, 41(2):271-284. https://doi.org/10.1109/TPAMI.2018.2799222

Shih Y, Krishnan D, Durand F, et al., 2015. Reflection removal using ghosting cues. Proc IEEE Conf on Computer Vision and Pattern Recognition, p.3193-3201. https://doi.org/10.1109/CVPR.2015.7298939

Solomon F, Ikeuchi K, 1996. Extracting the shape and roughness of specular lobe objects using four light photometric stereo. *IEEE Trans Patt Anal Mach Intell*, 18(4):449-454. https://doi.org/10.1109/34.491627

Sun C, Liu SC, Yang TT, et al., 2016. Automatic reflection removal using gradient intensity and motion cues. Proc 24th ACM Int Conf on Multimedia, p.466-470. http://doi.org/10.1145/2964284.2967264

Sunkavalli K, Zickler T, Pfister H, 2010. Visibility subspaces: uncalibrated photometric stereo with shadows. 11th European Conf on Computer Vision, p.251-264. https://doi.org/10.1007/978-3-642-15552-9_19

Tan RT, Ikeuchi K, 2008. Separating reflection components of textured surfaces using a single image. *IEEE Trans Patt Anal Mach Intell*, 27(2):353-384. https://10.1109/TPAMI.2005.36

Tankus A, Kiryati N, 2005. Photometric stereo under perspective projection. 10th IEEE Int Conf on Computer Vision, p.611-616. https://doi.org/10.1109/ICCV.2005.190

Wan RJ, Shi BX, Hwee TA, et al., 2016. Depth of field guided reflection removal. IEEE Int Conf on Image Processing, p.21-25. https://doi.org/10.1109/ICIP.2016.7532311

Wan RJ, Shi BX, Duan LY, et al., 2018. CRRN: multi-scale guided concurrent reflection removal network. Proc IEEE Conf on Computer Vision and Pattern Recognition, p.4777-4785.

Wang F, Ainouz S, Petitjean C, et al., 2017. Specularity removal: a global energy minimization approach based on polarization imaging. *Comput Vis Image Underst*, 158:31-39. https://doi.org/10.1016/j.cviu.2017.03.003

Woodham RJ, 1979. Photometric stereo: a reflectance map technique for determining surface orientation from image intensity. 22nd Annual Technical Symp on Image Understanding Systems and Industrial Applications I, p.136-143. https://doi.org/10.1117/12.956740

Wu L, Ganesh A, Shi B, et al., 2010. Robust photometric stereo via low-rank matrix completion and recovery. 10th Asian Conf on Computer Vision, p.703-717. https://doi.org/10.1007/978-3-642-19318-7_55

Yang J, Gong D, Liu LQ, et al., 2018. Seeing deeply and bidirectionally: a deep learning approach for single image reflection removal. Proc European Conf on Computer Vision, p.654-669. https://doi.org/10.1007/978-3-030-01219-9_40

Yang QX, Tang JH, Ahuja N, 2015. Efficient and robust specular highlight removal. *IEEE Trans Patt Anal Mach Intell*, 37(6):1304-1311. https://doi.org/10.1109/TPAMI.2014.2360402

Yeung SK, Wu TP, Tang CK, et al., 2015. Normal estimation of a transparent object using a video. *IEEE Trans Patt Anal Mach Intell*, 37(4):890-897. https://doi.org/10.1109/TPAMI.2014.2346195

Yu C, Seo Y, Lee SW, 2010. Photometric stereo from maximum feasible Lambertian reflections. 11th European Conf on Computer Vision, p.115-126. https://doi.org/10.1007/978-3-642-15561-1_9

Zhang XE, Ng R, Chen QF, 2018. Single image reflection separation with perceptual losses. Proc IEEE Conf on Computer Vision and Pattern Recognition, p.4786-4794. https://doi.org/10.1109/CVPR.2018.00503

Zhou ZL, Wu Z, Tan P, 2013. Multi-view photometric stereo with spatially varying isotropic materials. IEEE Conf on Computer Vision and Pattern Recognition, p.1482-1489. https://doi.org/10.1109/CVPR.2013.195