Photometric Stereo via Expectation Maximization

Tai-Pang Wu, Member, IEEE, and Chi-Keung Tang, Member, IEEE

Abstract—This paper presents a robust and automatic approach to photometric stereo, where the two main components, namely surface normals and visible surfaces, are respectively optimized by Expectation Maximization (EM). A dense set of input images is conveniently captured using a digital video camera while a handheld spotlight is being moved around the target object and a small mirror sphere. In our approach, the inherently complex optimization problem is simplified into a two-step optimization, where EM is employed in each step: 1) Using the dense input, the weight or importance of each observation is alternately optimized with the normal and albedo at each pixel and 2) using the optimized normals and employing the Markov Random Fields (MRFs), surface integrabilities and discontinuities are alternately optimized in visible surface reconstruction. Our mathematical derivation gives simple updating rules for the EM algorithms, leading to a stable, practical, and parameter-free implementation that is very robust even in the presence of complex geometry, shadows, highlight, and transparency. We present high-quality results on normal and visible surface reconstruction, where fine geometric details are automatically recovered by our method.

Index Terms—Photometric stereo, expectation maximization, normal, albedo and visible surface reconstruction.

1 INTRODUCTION

Two typical problems of shape from photometric stereo are: normal reconstruction from images and surface reconstruction from normals. This paper presents an approach that uses a set of dense measurements to robustly recover albedos and normals, and then uses the optimized normals to recover a high-quality visible surface that preserves the overall shape as well as fine geometric details. The optimization algorithms used in our approach are formulated using Expectation Maximization (EM).

Classical photometric stereo approaches used two views with known albedos [28] or three views [13] to solve the reflectance equation, and recover surface gradients of a Lambertian surface. Because the Lambertian assumption can be easily violated in the presence of shadows, highlights, and complex surface materials, approaches using more than three images were proposed [7], where non-Lambertian observations are discarded by consistency checking of all combinations; but it will become computationally expensive when the number of images increases significantly. Alternative approaches using model fitting have been proposed [20], [24], [18], [23], which impose relevant constraints to achieve more robust solutions. These methods are more capable in handling objects made of complex materials, but they can be vulnerable to numerical instability, because the reflectance models used are typically nonlinear and more complex than the Lambertian model. Usually, a few global and local parameters need to be estimated first to make the optimization less susceptible to local optimum.

Our photometric stereo approach exploits the data redundancy inherent in a dense image set to derive how well an observation fits the Lambertian model. The Lambertian reflectance model is simple, which is defined by the normal (unknown) and the light (estimated by a simple setup) at a pixel. The main issue lies in determining the degree an observation fits the Lambertian model. Our approach distinguishes Lambertian and Lambertian observations using soft labels (weight). Because the contributing weight of each observation as well as the normal at each pixel is unknown, alternating optimization approaches, such as EM, is particularly suitable to solve the optimization problem, which has also nice convergence properties [5]. Once we have obtained a set of optimized normals, the goal is to compute a high-quality visible surface, which should preserve the overall shape, surface smoothness, as well as surface orientation discontinuities. Noting that surface orientation discontinuity does not exist without smoothness in its vicinity, we also use EM to solve the optimization problem, by cooperating the optimization of surface integrabilities and orientation discontinuities in an alternating manner. As we will demonstrate, our EM approach to photometric stereo is very robust and produces very good reconstruction results in the presence of complex geometry, complex surface reflectance even transparency, shadows, and highlight.

In summary, we present a unifying framework to photometric stereo, where the two main components, namely surface normal and visible surface estimation, are, respectively, optimized by EM. We integrate in this paper the preliminary versions [30] and [31] by an overall objective function and provide more examples and analysis of the EM framework. The organization of the paper is as follows: Section 2 reviews related work in photometric stereo and visible surfaces reconstruction. Section 3 defines
our overall global objective function, where the complex optimization is simplified into a two-step optimization. Mathematical derivations of the two steps will be described in the following sections: First, using the noisy and dense input, Section 4 describes our EM algorithm in learning relevant observations, recovering surface normals and albedos, and inferring the optimal parameters so that no free user-supplied parameter is required. Given a set of normals estimated at all pixels, in Section 5, we present our EM optimization algorithm that alternately optimizes until convergence the surface integrabilities and discontinuities inherent in the normal field, in order to derive a segmented visible surface description of the target scene. Finally, we conclude the paper in Section 7.

2 RELATED WORK

2.1 Photometric Stereo

Woodham [28] first introduced photometric stereo for Lambertian surfaces in which two images are used to solve the reflectance equation for recovering surface gradients of a Lambertian surface. Three images are used when the albedo is unknown [13]. Over the past three decades, extensive research on more robust techniques for photometric stereo have been reported:

- **Four images.** Coleman and Jain [7] used four photometric images to compute four albedo values at each pixel, using the four combinations derived from any three of the given images. In the presence of specular highlight, the computed albedos may not be identical, indicating that some measurement must be excluded. In [23], four images were also used. Barsky and Petrou [3] showed that [7] is still problematic if shadows are present, and generalized [7] to handle color images. These methods are sensitive to noise caused by violations to the Lambertian model.

- **Reference objects.** In [14], a reference object was used to perform photometric stereo reconstruction in which isotropic materials were assumed. In this approach, the outgoing radiance functions for all directions are tabulated to obtain an empirical reflectance model. In [12], a similar technique was used to compute surface orientations and reflectance properties, where the authors made use of the orientation consistency proposed to establish the correspondence between an unknown object and a known reference object. In many cases, however, a reference object for establishing correspondence is unavailable. A simplified reflectance model will then be used.

- **Reflectance models.** By considering diffuse and non-Lambertian surfaces, Tagare and deFigueiredo [24] proposed the $m$-lobed reflective map to solve the problem. Kay and Caelly [18] extended [24], by proposing the use of nonlinear regression to a larger number of input images. Solomon and Ikeuchi [23] separated the object into different areas and used the Torrance-Sparrow model to compute the surface roughness. Nayar et al. [20] used a hybrid reflectance model (Torrance-Sparrow and Beckmann-Spizzichino) and recovered not only the surface normals but also the parameters of the reflectance model. Georghiades [10] incorporated the Torrance-Sparrow model in his photometric stereo algorithm. In these approaches, the models used are usually quite complex, and a larger number of parameters are estimated. Basri and Jacobs [4] showed that

the set of images of a convex Lambertian object can be approximated by low-dimensional linear subspaces. They assumed isotropic and distant light sources whose directions may be unknown or arbitrary. Shape recovery is performed in a low-dimensional space. Goldman et al. [11] proposed a photometric stereo method using the Ward model to recover the shape (normals) and BRDFs using an alternating optimization scheme. Unlike their earlier work [12], a reference object is not needed, which is solved as part of the reconstruction process. Since they only used a sparse set of samples, the light calibration should be quite accurate, and severe highlight and cast shadows must not be present unless an outlier detection procedure or robust estimator is used. In [6], photometric stereo reconstruction for glossy surfaces was proposed, where the Ward model is also used.

**Markov random fields.** Photometric stereo was, respectively, solved using tensor belief propagation [25] and graph cuts [29]. A dense set of images with variable illumination is captured at a fixed viewpoint. Given the data, the normal at each pixel is optimized by the respective Markov Random Fields (MRFs) algorithms. The initial normal at each pixel is obtained by solving a linear system of equations, which is obtained by dividing each image by a chosen image called denominator image. Although the MRF approach produces some of the best normal reconstruction results, the approach suffers from the following problems:

- Assuming the Lambertian model, the albedo is canceled out in the equations after dividing each image with the denominator image.
- Least-square plane fitting is incapable of rejecting non-Lambertian observations. Outliers significantly affect the fitting result.
- If the denominator image contains a pixel observation that is non-Lambertian, the whole set of linear equations corresponding to that pixel will produce unpredictable results because of the violation of Lambertian assumption.

The noise due to non-Lambertian observations that cannot be handled by least-squares fitting was addressed by the MRF refinement process. Although discontinuity-preserving compatibility functions were used, the MRF smoothing effect was applied globally because no prior knowledge is available on which regions should be smoothed more or otherwise. As in other MRF algorithms, a free parameter should be supplied by the user to control the degree of normal smoothness, whose value is empirically obtained that varies with different scenes. In [27], outlier handling was addressed in photometric stereo, where the inlier and outlier processes were modeled using MRF. In this approach, each scaled pixel normal is modeled as a continuous hidden variable whose distribution is approximated by a Gaussian. To address the inherent computational intractability, the mean-field EM algorithm is used.

2.2 Visible Surfaces from Surface Gradients

Many techniques have been proposed to derive a segmented visible surface description, given a dense and possibly imperfect set of normals that can be obtained from photometric stereo or shape from shading. Traditional approach enforces the surface integrability or the zero-curl constraint across the whole normal field [8] so that the height at any given point can be computed by summing up
they used edge detection to label discontinuities via the ability to preserve manually labeled discontinuities. Later, in the image domain. Recently, a kernel-based uniform apply the surface integrability constraint everywhere. In these cases, it is inappropriate to produce highly detailed normal fields, where surface normal reconstruction methods [11], [29], [25], [30] can result (e.g., by [9] and [19]). However, many state-of-the-art distortion can be produced even for some simple examples without basis functions is proposed by Frankot and Chellappa [9]. They projected the surface gradients onto a set of integrable basis functions to enforce surface integrability by using Fourier basis functions. Variants such as wavelet basis functions were proposed [15]. Karacali and Snyder [16] made use of an orthonormal set of gradient fields and projected the measured gradient field onto a feasible subspace to generate an integrable surface with surface gradients resembling the measured ones. Shape from shapenet was proposed in [19]. This method correlates the measured surface gradients with the gradients of a bank of shapenet basis functions that satisfy a set of constraints such as admissibility, minimal ambiguity, uniform coverage, and preservation of phase information. One drawback common to basis function-based methods is that severe surface distortion can be produced even for some simple examples with piecewise smooth fields (see Fig. 8).

Most surface-from-gradients methods globally enforce the integrability constraint in the image domain. This is, however, fundamentally problematic because a continuous surface may not be differentiable everywhere. When these methods are applied to a partially integrable field, normal orientations must inevitably be altered in order to satisfy the zero-curl constraint, which may lead to a distorted surface result (e.g., by [9] and [19]). However, many state-of-the-art normal reconstruction methods [11], [29], [25], [30] can produce highly detailed normal fields, where surface discontinuities or abrupt changes in normal orientations should be preserved. In these cases, it is inappropriate to uniformly apply the surface integrability constraint everywhere in the image domain. Recently, a kernel-based approach was used in [16], which enforces partial integrability to preserve manually labeled discontinuities. Later, they used edge detection to label discontinuities via multiscale analysis [17]. The examples presented in [17] are planar or simple curved surface, where only one result on a simple real image was demonstrated.

3 THE OVERALL OBJECTIVE FUNCTION
We formulate in this section the overall objective function to address the photometric stereo problem, where dense and noisy measurements are used. Due to the complexity of the optimization problem where a substantial number of parameters need to be estimated, we propose a two-step optimization by EM: Normals and albedos are optimized in the first step (photometric stereo). Using the optimized normals, in the second step (visible surface reconstruction), a surface is inferred that preserves discontinuities inherent in the normal field.

Given a set of observations \( O \), which is directly derived from the dense set of captured images, our goal is to use EM to optimize a set of parameters (that is, normals, albedos, weight of each observation, integrability, and discontinuity). We let \( \Theta \) and \( \lambda \) be the respective set of parameters to be optimized in step 1 and step 2. Typical of EM algorithms, we use a set of hidden variables, respectively denoted by \( R \) and \( S \), to formulate the corresponding EM algorithms of the two steps. In our formulation, \( R \) is regarded as the Lambertian sample flag and \( S \) is used as the connectivity flag.

The overall objective function can, therefore, be written as

\[
(\Theta^*, \lambda^*) = \arg\max_{\Theta, \lambda} P(\Theta, \lambda, R, S | O).
\]

Equation (1) can be decomposed into the following:

\[
\begin{align*}
\arg\max_{\Theta, \lambda} P(\Theta, \lambda, R, S | O) \\
= \arg\max_{\Theta, \lambda} P(\Theta, \lambda, R, S | O)P(S | O) \\
= \arg\max_{\Theta, \lambda} P(\lambda, R, \Theta, S, O)P(\Theta | S, O)P(S | O) \\
= \arg\max_{\Theta, \lambda} P(\lambda, R, \Theta, \tilde{O})P(S, O | \Theta)P(\Theta | S, O)P(S | O).
\end{align*}
\]

We simplify the complicated problem by first solving the optimal \( \Theta^* \), which is used to define \( \tilde{O} \). Then, (2) is simplified into a two-step optimization:

\[
\Theta^* = \arg\max_{\Theta} P(S, O | \Theta)P(\Theta)P(S | O),
\]

\[
\lambda^* = \arg\max_{\lambda} P(R, \tilde{O} | \lambda)P(\lambda)P(R | \tilde{O}).
\]

Typical of Bayesian optimization, \( P(\Theta) \) and \( P(S | O) \) in (3) and \( P(\lambda) \) and \( P(R | \tilde{O}) \) in (4) are assumed to be uniform distributions because we do not have any information for their estimation.

In the following, Section 4 describes an EM algorithm to solve (3). The EM algorithm computes the optimal normals and albedos. Section 5 describes an EM algorithm to solve (4). This EM algorithm refines the optimized normals and labels all discontinuities inherent in the normal fields. As we shall see, both algorithms give simple EM updating rules, which lead to a simple implementation where user-supplied parameters are not required because all parameters are automatically optimized in the EM framework.
In this section, we formulate an EM algorithm to solve $\Theta^*$, which automatically identifies relevant Lambertian observations, by exploiting the data redundancy inherent in the dense and noisy data. Our normal results demonstrate significant improvement over [32] while no MRF smoothing refinement is performed on the normal field and there is no parameter to set.

Our EM algorithm estimates the surface albedos and normals from a set of dense and noisy measurements. The input consists of a dense set of noisy photometric images conveniently captured using a digital video camera, a reflective mirror sphere, and a handheld spotlight, which is the same as that used in [32] (shown in Fig. 1), where least-square plane fitting was used for initial normal estimation. Note that no special handling was performed in [32] for unreliable data or outliers generated by non-Lambertian phenomena. In practice, however, these observations occupy a nonnegligible portion in the captured data, due to the restrictive Lambertian model and the complexity of real surface geometry and material.

Now, let us suppose that the measurement error for each observation is known. We could have performed weighted least-square plane fitting to weaken the contribution of defective data. However, given the simple data capture system, it is very difficult to estimate such measurement errors. Instead of building an expensive and accurate capturing system, we propose a data-driven approach to estimate the weight of each observation by utilizing useful information inherent in the dense but noisy image set.

4.1 Overview

While still using the Lambertian model instead of canceling out the albedo to simplify computation as in [32], we use the albedo as one of the contributing factors in estimating the weight of each observed intensity. The idea is as follows: Consider a pixel location $i$. Suppose that the albedo $\rho_i$ is known, given a captured image $I_i$, the observed pixel intensity $I_{it}$ with the corresponding light direction $L_t$, we model the probability of the intensity $I_{it}$ generated by the Lambertian model without shadow and specular highlight to be negatively exponential to

$$||I_{it} - \rho_i n_i \cdot L_t||,$$

where $n_i$ is the normal at pixel $i$. Thus, if the albedo is known, more information concerning the observations can be extracted. However, the derivation of albedo itself is a difficult problem. In this paper, we demonstrate how the weight or importance of each sample can be optimized alternately with the estimation of albedo and surface normal, using an EM framework to obtain accurate results.

While we argue that least-square plane fitting without a proper contribution weight for each observation is not a good solution, some useful lesson can still be learned from [32]. Suppose that each image is a candidate denominator image to generate a ratio image, where each pixel in a ratio image is expressed by $I_t = \frac{n_i L_t}{n_i n_t}$, where $I_t$, called the denominator image, is an image selected from the dense set of captured images. By using no less than three ratio images, a local estimation of the normal is obtained at each pixel: $A_t I_{dx} + B_t I_{dy} + C_t = 0$, where $A_t = I_{dx} - I_{d_{x,z}}, B_t = I_{dy} - I_{d_{y,z}}, C_t = I_{dz} - I_{d_{z,z}}, L_t = [l_{x,z}, l_{y,z}, l_{z,z}]^T$ is the light direction for image $t = 1 \cdots T$, and $n_i = [x, y, z]^T$ is the unknown normal to be estimated. Note that this local estimation is valid only when the Lambertian assumption holds.

Given a total of $T$ different observations for a pixel location, we can produce $T$ different planes by using all images successively as the denominator. For the subset of denominators consisting of non-Lambertian observations, the orientations of the fitted planes are arbitrary without any correlation. For the other subset of denominators whose observations are explained by the Lambertian model, however, the orientations of the resulting planes should cluster themselves together. Despite that such estimated planes are not error-free because of the presence of outliers, the cluster does constrain the solution space for the optimal surface orientation at the pixel.

4.2 The Objective Function

In this section, the index of pixel locations will be dropped for simplicity of notations. The EM algorithm to be described is applied independently at each pixel location. No smoothing or MRF refinement is performed.

We define $O = \{o_t\}$ to be the set of observations, where $t = 1 \cdots T$ and $T$ is the total number of captured images; $o_t = \{I_t, n_t\}$, where $I_t$ (a 3-vector in RGB space) is the observed pixel intensity for image $t$ and $n_t$ is the normal obtained after plane fitting with image $t$ as the denominator image.

To encode the distribution of $\{n_t\}$, a $3 \times 3$ covariance matrix $K$ is used to store the second-order moment collection to represent orientation distribution. The normal is the direction given by the largest variance in $K$.

Our goal is to find the optimal albedo $\rho$ (a 3-vector in RGB space) and covariance matrix $K$ given the pixel observations. From (3), we want to estimate the following:

$$\Theta^* = \arg \max_{\Theta} \mathbb{P}(O, S | \Theta),$$

where $I_t$ is the intensity at pixel $t$, $n_t$ is the normal at pixel $t$, and $L_t$ is the light direction at pixel $t$. The denominator image is a ratio image generated by the Lambertian model without shadow and specular highlight. In practice, each pixel $i$ has $T$ observations with intensity in the 50th percentile or higher, because dark pixels tend to be affected by shadows. These brighter samples provide sufficient redundancy for robust estimation. We could have used $T$ observations to perform the computation. This will produce similar results, because non-Lambertian observations have been attenuated after running the EM algorithm. Since we already have a dense set of images, it is fine to remove known and contaminated pixels to speed up processing and reduce memory requirement. Our technique itself is meant to handle all data. We have empirically tested and confirmed that similar results are obtained while the execution time is longer when the entire collection of $T$ observations is considered.
where \( P(O, S | \Theta) \) is the complete-data likelihood we want to maximize. \( \Theta = \{ K, \rho, \alpha, \sigma \} \) is a set of parameters to be estimated, and \( S = \{ s_t \} \) is a set of hidden states indicating which observation is generated by the Lambertian model: \( s_t = 1 \) if \( o_t \) is generated by Lambertian model, \( s_t = 0 \) otherwise. \( \alpha \) and \( \sigma \) are, respectively, the proportion of Lambertian observations and the standard deviation of (5), which are the parameters to help us to find the solution and will be described shortly.

Our EM algorithm estimates (6) by finding the expected value of the complete-data log-likelihood \( \log P(O, S | \Theta) \) w.r.t. \( S \) given the observation \( O \) and the current estimated parameters:

\[
Q_\Theta(\Theta') = \sum_{S \in \varphi} \log P(O, S | \Theta)P(S | O, \Theta')dS, \tag{7}
\]

where \( \Theta' \) are parameters at the current iteration and \( \varphi \) is a space containing all \( S \) of size \( T \).

### 4.3 Expectation

In this section, we address how to estimate the marginal distribution \( p(s_t | o_t, \Theta') \) so that we can maximize the expectation \( Q_\Theta \) defined by (7) by proceeding to the next iteration, given the current values of the parameters.

If \( s_t \) is known, the observation \( o_t \) generated by the Lambertian model minimizes (5) and \( n_t^T K^{-1} n_t \). Suppose that the noise distribution of (5) and the jittering distribution of \( n_t \) are Gaussian distributions, and the existence of non-Lambertian observations follow a uniform distribution, then the observation probability of \( o_t \) is:

\[
p(o_t | s_t, \Theta') \propto \begin{cases} 
\exp \left( -\frac{||I_t - \rho n_t \cdot L_t||^2}{2\sigma^2} \right) \exp \left( -\frac{1}{2} n_t^T K^{-1} n_t \right), & \text{if } s_t = 1; \\
\frac{1}{C}, & \text{if } s_t = 0. 
\end{cases} \tag{8}
\]

Based on the uniform distribution assumption, the choice of \( C \) should be \( \max \| I_t - \rho n_t \cdot L_t \| \). However, in practice, the assumption can be violated seriously. To lower the chance of wrong classification, we choose \( C = C_m = \text{mean} \| I_t - \rho n_t \cdot L_t \| \) because smaller \( C \) tends to classify more observations to \( s_t = 0 \) (non-Lambertian). This lowers the probability of including non-Lambertian samples, while we still have sufficient redundancy for estimation robustness. To calculate \( C \), we choose \( \rho \) to be the color that has the median gray-level intensity. Indeed, \( C \) needs not to be precise. In all of our experiments, varying \( C, C_m \leq C \leq 2C_m \) produces very similar results, and thus, this constant is not critical.

Let \( \alpha \) be the proportion of observations generated by the Lambertian model. Then we have a mixture probability of the observations:

\[
p(s_t = 1) = \alpha. \tag{9}
\]

So, given \( \Theta' \) only, we have

\[
p(o_t | \Theta') \propto \alpha \exp \left( -\frac{||I_t - \rho n_t \cdot L_t||^2}{2\sigma^2} \right) \exp \left( -\frac{1}{2} n_t^T K^{-1} n_t \right) + \frac{1 - \alpha}{C}. \tag{10}
\]

Let \( w_t \) be the probability of \( o_t \) generated by the Lambertian model. Then,

\[
w_t = p(s_t = 1 | o_t, \Theta') = \frac{p(o_t, s_t = 1 | \Theta')}{p(o_t | \Theta')}
= \frac{\alpha \exp \left( -\frac{||I_t - \rho n_t \cdot L_t||^2}{2\sigma^2} \right) \exp \left( -\frac{1}{2} n_t^T K^{-1} n_t \right)}{\alpha \exp \left( -\frac{||I_t - \rho n_t \cdot L_t||^2}{2\sigma^2} \right) \exp \left( -\frac{1}{2} n_t^T K^{-1} n_t \right) + \frac{1 - \alpha}{C}}. \tag{11}
\]

Hence, in the E-step of our EM algorithm, we compute \( w_t \) for all \( t = 1 \cdots T \).

#### 4.4 Maximization

In this section, we maximize the likelihood (6) given the marginal distribution \( w_t \) computed in the E-Step.

Since we only have two states \( \{0, 1\} \) for each \( s_t \), the \( Q_\Theta \) function (7) is

\[
Q_\Theta(\Theta, \Theta') = \sum_t \log p(o_t, s_t = 1 | \Theta) w_t
+ \sum_t \log p(o_t, s_t = 0 | \Theta)(1 - w_t)
= \sum_t \log \left( \frac{\alpha}{\sigma \sqrt{2\pi}} \exp \left( -\frac{||I_t - \rho n_t \cdot L_t||^2}{2\sigma^2} \right) \right) w_t
+ \sum_t \log \left( \frac{1 - \alpha}{C} \right) \exp \left( -\frac{1}{2} n_t^T K^{-1} n_t \right) w_t
+ \sum_t \log \left( \frac{1}{C} \right) \exp \left( -\frac{1}{2} n_t^T K^{-1} n_t \right) (1 - w_t). \tag{12}
\]

To maximize (12), we set the first derivative of \( Q \) w.r.t. \( \alpha \), \( \sigma \), \( \rho \), and \( K \), respectively, equal to zero and obtain the following simple parameter updating rule for \( \Theta \), and thus, the M-step of our EM algorithm:

\[
\alpha = \frac{1}{T} \sum_t \omega_t,
\sigma = \frac{\sum_t ||I_t - \rho n_t \cdot L_t||^2 \omega_t}{\omega_t},
\rho = \frac{1}{\sum_t (n_t \cdot L_t)^2 \omega_t} \sum_t I_t (n_t \cdot L_t) \omega_t,
K = \frac{1}{\sum_t \omega_t} \sum_t n_t n_t^T \omega_t. \tag{13}
\]

The E-Step and M-Step are executed alternately until the process converges. While the convergence of EM was well established [5], we show in Fig. 2 the error curves of selected pixels for the example Rope. Note that although the error curves, which depend on the corresponding initial guesses and the error surfaces, are different for the sampled pixels, the demonstrated monotonicity verifies the convergence of the EM algorithm.
Upon convergence, we apply eigendecomposition on $K$ to obtain the optimal normal direction. The eigenvector corresponding to the largest eigenvalue gives the normal direction.

Note that, using our method, we produce not only surface normals but also surface albedo $\rho$, and the weights $w_i$, indicating the degree an observation $o_i$ is consistent with the Lambertian model. Some selected frames are shown in Fig. 3, where the Lambertian and non-Lambertian components were produced by multiplying the input image by $w_i$ and $1 - w_i$, respectively.

5 Integrability and Discontinuity Optimization by Expectation Maximization

Now, given $\Theta^*$, we can solve (4). That is, given the normal map estimated using the above EM algorithm, we estimate the discontinuity weights, which will be utilized for computing a smooth height field wherein fine details and surface discontinuities are preserved.

Observe that surface orientation discontinuities do not exist without smoothness in their vicinity: Typical examples are intersections between two smooth surfaces and depth discontinuities resulting from one smooth surface occluding the other. So, the EM framework is particularly suitable in this subproblem, which requires cooperating surface integrability and discontinuity in an alternating optimization manner. In the following, we define the terms of our objective function that form the basis of our EM solution to solve for the discontinuity map, given a dense normal field.

5.1 The Objective Function

Let $i$ be the location index of the normal map, where $i \in \{1, \ldots, N\}$ and $N$ is the size of the normal map. We let $L$, $R$, $T$, and $B$ be the respective pixel location indices of the left, right, top, and bottom neighbors of pixel $i$.

We define $\tilde{O} = \{o_i = (\tilde{p}_i, \tilde{q}_i)\}$ to be the set of observations, where $\tilde{p}_i$ and $\tilde{q}_i$ are the partial derivatives at location $i$ in the $x$ and $y$ directions, respectively, which are kept constant as observations. They are obtained from

the normal map derived from the optimized $\Theta^*$, that is, by expressing $n$ as $[p \ q \ 1]^T$.

Given $\tilde{O}$, our goal is to estimate the optimal partial derivatives $p_i$ and $q_i$. Following (4):

$$\lambda^* = \arg \max_\lambda P(R, \tilde{O} | \lambda),$$

where $P(R, \tilde{O} | \lambda)$ is the complete-data likelihood we want to maximize, $\lambda = \{\{p_i, q_i\}, \sigma_1, \sigma_2\}$ is the set of parameters to be estimated, $\sigma_1$ and $\sigma_2$ are the parameters of certain distributions to be introduced, and $R = \{r_i\}$ is a set of hidden variables on the connectivity labels. Suppose that $i$ and $j$ are first-order neighbors, they are connected if $r_i = r_j$. Otherwise, $r_i \neq r_j$ when there exists a salient discontinuity between $i$ and $j$.

To estimate (14), the EM algorithm computes the expected value of the complete-data log-likelihood $\log P(\tilde{O}, R | \lambda)$ w.r.t. $R$ given the observation $\tilde{O}$ and the current estimated parameters $\lambda^*$:

$$Q(\lambda, \lambda^*) = \sum_{R \in \varphi} \log P(\tilde{O}, R | \lambda)P(R | \tilde{O}, \lambda^*),$$

where $\varphi$ is a space containing all $R$ of size $N$.

Here, the situation differs from normal reconstruction in the previous section, where dense data per pixel are available and MRF smoothing across pixels is unnecessary and inadvisable. While there is just one optimized normal per pixel available, for reconstructing a continuous surface, appropriate connectivity information on the image grid should be considered. In this subproblem, we make use of the MRF to model the connectivity and the dependency of the neighboring pixels in the complex labeling problem.

We first review here two standard MRF assumptions [5]:

1. The hidden variable $r_i$ depends only on the hidden variables of its first-order neighbors.  
2. The observation at $i$ depends only on the hidden variable at $i$.  

Fig. 2. Number of EM iterations and convergence plots of randomly selected pixels for the example Rope. The $y$-axis represents the value of (12). The $x$-axis is the number of EM iterations.

Fig. 3. Two inputs and the corresponding Lambertian and non-Lambertian components of Toy car are shown. While shadows and highlights are correctly classified, some Lambertian samples are classified into the non-Lambertian set due to the imprecise $C$ (discussed in the text). However, given these correctly classified Lambertian samples, a decent estimation of the normal map and surface albedo can still be achieved.
5.2 Expectation
In this section, we describe how to estimate \( P(R | \hat{O}, \lambda') \) so that we can maximize the expectation \( Q_\lambda \) defined by (15) in each iteration given the current estimated parameters.

Without lost of generality, the marginal probability \( p(r_i | o_i, \lambda') \) is always equal to 1 since there is no connectivity consideration. Since we want to locate discontinuities, we need to define \( p(r_i, r_j | o_i, o_j, \lambda') \) (or \( p(r_i, r_j | \lambda') \)) because of MRF assumption (1), where \( i \) and \( j \) are the indices of first-order neighbors. To know whether discontinuity exists between \( i \) and \( j \), we use the measurement of similarity of neighbors (16) and integrability (18).

Neighborhood similarity. We introduce neighborhood inference by minimizing the following equation, which is related to the second-order derivatives, and in other words, the curvature of the underlying surface:

\[
\|p_i - p_R\| + \|q_i - q_B\|. \tag{16}
\]

Integrability. Given a 2D surface normal map, our goal is to recover a segmented height field, where fine details on smooth surfaces as well as surface discontinuities should be preserved. The height field is represented as a set of tuples \((i, f_i)\), where \( f_i \) is the height at \( i \). A normal at \( i \) is given by:

\[
n_i = \frac{1}{\sqrt{\frac{\partial f_i}{\partial x}^2 + \frac{\partial f_i}{\partial y}^2}} \left( \frac{\partial f_i}{\partial x}, \frac{\partial f_i}{\partial y} \right)^T.
\]

Note that a normal, as obtained in (13), can be written in this form.

The integrability (or zero-curl) constraint allows the surface (height field) to be estimated by integration so that the recovered surface does not depend on the choice of the integration path [8]. The constraint can be implemented by minimizing the following:

\[
\left\| \frac{\partial f_i}{\partial x} - \frac{\partial f_j}{\partial y} \right\| \tag{17}
\]

Let \( p_i = \frac{\partial f_i}{\partial x} \) and \( q_i = \frac{\partial f_i}{\partial y} \) be the partial derivatives in the \( x \) and \( y \) directions, respectively. The discrete version of (17) is

\[
\left\| (p_i - p_R) + (q_R - q_i) \right\|. \tag{18}
\]

Our goal is to find a configuration for the weighted discontinuity map that minimizes (18).

However, (18) is difficult to be embedded into an MRF network together with the neighborhood inference given by (16). So, instead of using (18) directly, by triangle inequality, we consider its upper bound as the approximation:

\[
\|p_i - p_R\| + \|q_i - q_B\|. \tag{19}
\]

Note that it is possible that the upper bound formulation would make the estimated surface normals strictly non-integrable (if the underlying normal field is integrable). As we shall see, we still have (28), whose solution is given by solving a Poisson equation, as the final bootstrap. The error is distributed to all pixels by considering all possible residuals on the image grid. A similar method is applied in [11] to minimize the zero-curl error of the estimated, nonintegrable normal map, in order to obtain a continuous surface.

In summary, if the surface connecting locations \( i \) and \( j \) is smooth (i.e., \( r_i = r_j \)), (16) and (19) should be minimized. To achieve this, for each \( p_i \) (respectively, \( q_i \)), both the right and bottom neighbors are considered. Because each \( i \) should be covered exactly once, without loss of generality, these two equations can now be written nicely into a single equation by replacing \( R \) or \( B \) with a neighborhood index \( j \).

Let \( v_{ij} = p(r_i = r_j | \lambda') \). Therefore, the marginal probability of the connectivity between \( i \) and \( j \), which considers both neighborhood similarity and integrability, turns out to have a simple form:

\[
v_{ij} = p(r_i = r_j | \lambda') \propto \exp \left( -\frac{\| p_i - p_j \|^2 + \| q_i - q_j \|^2}{2\sigma_1^2} \right). \tag{20}
\]

So, in the E-Step, we solve \( v_{ij} \) for every first-order pair of neighboring pixels \( \{i, j\} \). Note that \( v \) encodes the discontinuity probability for every pair of pixels, thus giving the weighted discontinuity map we desire.

5.3 Maximization
We decompose \( P(\hat{O}, R | \lambda) \) by first rewriting it into a combination of simple elements. Based on the two MRF assumptions, the likelihood can be broken down in this way:

\[
P(\hat{O}, R | \lambda) = \prod_{R \in \varphi} \prod_{i \in \mathcal{N}_i} p(r_i | r_j, \lambda) p(o_i | r_i, \lambda), \tag{21}
\]

where \( \mathcal{N}_i \) is a set of right and bottom neighbors of \( i \). \( p(r_i | r_j, \lambda) = 1 \) if \( j \) does not exist.

Now, the \( Q_\lambda \) function in (15) can be rewritten as

\[
Q_\lambda(\lambda, \lambda') = \sum_{R \in \varphi} \log \left( \prod_{i \in \mathcal{N}_i} \prod_{j \in \mathcal{N}_j} p(r_i | r_j, \lambda) p(o_i | r_i, \lambda) \right) P(R | \hat{O}, \lambda')
\]

\[
\quad = \sum_{R \in \varphi} \sum_{i \in \mathcal{N}_i} \sum_{j \in \mathcal{N}_j} \log(p(r_i | r_j, \lambda)) P(R | \hat{O}, \lambda') + \sum_{R \in \varphi} \sum_{i \in \mathcal{N}_i} \log(p(o_i | r_i, \lambda)) P(R | \hat{O}, \lambda').
\]

(22)

Now, we model the likelihood \( p(o_i | r_i, \lambda) \) by minimizing the difference between \( o_i \) and the estimated output. So, we have

\[
p(o_i | r_i, \lambda) \propto \exp \left( -\frac{\| p_i - \hat{p}_i \|^2 + \| q_i - \hat{q}_i \|^2}{2\sigma_2^2} \right), \tag{23}
\]

where \( \sigma_2 \) is the standard deviation of the difference between the observation and the estimated output.

Since we only have two states \( \mathcal{G}_i = \{ r_i = r_j, r_i \neq r_j \} \), (22) can be further simplified:

\[
Q_\lambda(\lambda, \lambda') = \sum_{R \in \varphi} \sum_{i \in \mathcal{N}_i} \sum_{j \in \mathcal{N}_j} \log(p(g | r_j, \lambda)) p(g | \lambda')
\]

\[
+ \sum_{i} \log(p(o_i | r_i, \lambda)) p(r_i | o_i, \lambda)
\]

\[
\quad = \sum_{i \in \mathcal{N}_i} \log \left( \frac{1}{\sigma_1 \sqrt{2\pi}} \exp \left( -\frac{k(i, j)}{2\sigma_1^2} \right) \right) v_{ij}
\]

\[
+ \sum_{i} \log \left( \frac{1}{\sigma_2 \sqrt{2\pi}} \exp \left( -\frac{m(i)}{2\sigma_2^2} \right) \right).
\]

(24)
where \( k(i, j) = \|p_i - p_j\|^2 + \|q_i - q_j\|^2 \) and \( m(i) = \|p_i - \bar{p}_i\|^2 + \|q_i - \bar{q}_i\|^2 \). Recall that \( p(r_i, \alpha, \lambda') = 1 \).

To maximize (24), we set the first derivative of \( Q \) w.r.t. \( \sigma_1, \sigma_2, p_i, \) and \( q_i \), respectively, equal to zero, and finally, obtain the parameter updating rules for \( \lambda \) as:

\[
\begin{align*}
\sigma_1 &= \frac{\sum_{(i,j) \in \mathcal{H}} k(i, j) v_{ij}}{\sum_{(i,j) \in \mathcal{H}} v_{ij}}, \\
\sigma_2 &= \frac{1}{N} \sum_i m(i), \\
p_i &= \frac{\sum_{j \in \mathcal{K}_i} p_j v_{ij} + \frac{\sigma_3}{\sigma_2} \bar{p}_i}{\sum_{j \in \mathcal{K}_i} v_{ij} + \frac{\sigma_3}{\sigma_2}}, \\
q_i &= \frac{\sum_{j \in \mathcal{K}_i} q_j v_{ij} + \frac{\sigma_3}{\sigma_2} \bar{q}_i}{\sum_{j \in \mathcal{K}_i} v_{ij} + \frac{\sigma_3}{\sigma_2}}, \\
\end{align*}
\]

(25)

where \( \mathcal{H} \) is a set of neighboring pixel pairs and \( \mathcal{K}_i \) is the set of first-order neighbors (top, bottom, left, and right) of \( i \).

Hence, in the M-Step, the updating rule (25) is applied to estimate a new set of parameter values. The E-Step and M-Step are iterated alternately until convergence. The marginal probability \( v_{ij} \) computed using the converged parameters (20) gives the weighted discontinuity map. Table 1 lists the number of iterations required for each example.

### 5.4 Height Field Estimation

Given the discontinuity map \( v \), let \( h_i \) be the height at location \( i \). Given a first-order neighbor pair \( i \) and \( j \), the residual of the height \( h_i \) w.r.t. \( h_j \) is defined by the difference between the estimated \( h_i \) and the height integrated starting from \( j \). With the discontinuity map, the residual is weighted by \( v_{ij} \). Mathematically, the residual is defined as follows:

\[
\text{Residual} = v_{ij} (h_i - h_j)
\]

The experiments were run on a shared CPU server with 4 Opteron (TM)
where the $p$s and $q$s are the output of the EM algorithm.
Thus, the overall residual $E$ of the reconstructed surface is defined by the summation of all the residuals in (26) for each pair of $i$ and $j$:

$$E(h) = \sum_{i} \left( v_{R}(h_{i} - h_{R} - p_{R})^{2} + v_{B}(h_{i} - h_{B} - q_{B})^{2} \right). \quad (27)$$

Since each residual is a convex function, the summation of the residuals, i.e., $E(h)$, is also a convex function making the optimization guarantee to converge. We can use a standard optimization package to minimize (27) to extract the height field. In our implementation, we solve $h_{i}$ by applying the iterative Gauss-Seidel method. By setting $\partial E(h)/\partial h_{i} = 0$, we have
In each iteration, for each $i$, we estimate $h_i$ by solving (28) until the algorithm converges or the maximum number of iterations is reached.

6 EXPERIMENTAL RESULTS

This section presents the results of normal and albedo estimation, followed by height estimation given the estimated normal map. One of the key contributions of this paper consists of the identification of non-Lambertian observations from dense and noisy photometric images. Without proper weighting of the data, non-Lambertian observations will seriously hamper the estimation accuracy, especially when least-squares fitting is used. Using the same dense data set *Teapot*, we first demonstrate the efficacy of our EM-based photometric stereo, by comparing with the classical Woodham’s method [28], [13] and the ratio-image approach (that is, plane fitting after albedo cancellation) in [32], where the least-squares fitting is used for normal estimation in both methods. See also Fig. 4, where we study in detail the *Teapot*’s surface. Note the complex geometry and texture of the object. The selected close-up views of the teapot reveal fine surface details and subtle geometry.

Refer to Fig. 5 for the normal and surface results generated using the two previous methods. While all the three approaches use the Lambertian model, we will show later (in the top row of Figs. 7 and 12) that our normal and surface results, which take into account the contributing weight of each observation estimated by EM, show the importance of soft labeling. Note that the surface generated by Woodham’s method [28], [13] is too flat (because of the strong ambient effect inherent in this data set), while the normal map generated by the ratio-image method contains artifacts due to the complex reflectance and texture of the real teapot. An interesting point to note is that although adversely affected by ambient illumination, the surface generated by Woodham’s method exhibits little artifact due to specularities, in contrast to results produced by many existing state-of-the-art techniques. This is due to the use of a large number of input images, which helps to ameliorate some of the strong specular effects.

For non-Lambertian model, we expect that approaches such as [11] would benefit from similar soft labeling of relevant observations, with one caveat: Due to the larger number of model parameters, a more general form of EM, that is, alternating optimization (AO), should be used. This will be our future work.

6.1 Normal and Albedo

We first demonstrate the improvement by comparing our method with [32] using the same input data. The synthetic case we use is *Three Spheres* and the real examples are *Teapot, Rope, Toy Car, Face*, and *Transparency*. The running times of chosen examples are tabulated in Table 2.

*Three spheres*. Figs. 6a and 6b show two input synthetic images of *Three Spheres*. The rendered image is generated by the Phong illumination model. Figs. 6e and 6f show, respectively, the normal map produced by [32] and our EM method. Note that they are shaded using the Lambertian model $(n \cdot L)$ for visualizing the accuracy of the recovered normals. The ground truth is shown in Fig. 6c.

Our estimated albedo is shown in Fig. 6d. Qualitatively, the appearance of Figs. 6c, 6e, and 6f is very similar. On the other hand, the difference-images show the improvement of our method in terms of accuracy. Fig. 6g is the remapped difference-image when comparing Figs. 6c and 6e, and Fig. 6h is the remapped difference-image when comparing Figs. 6c and 6f. Note the presence of three halos in Fig. 6g while Fig. 6h is nearly totally white. We measure the mean angular error of the recovered normals to evaluate both methods quantitatively. Using *Three Spheres*, the mean error of the result produced by [32] is 4.041 degrees while the error of our EM result is only 1.5065 degrees.

**Real examples.** Our EM method shows very significant improvement in the presence of a large amount of noises in *Teapot* case when compared with [29], [32]. The top row of Fig. 7 shows the input and result of the *Teapot*. In Fig. 7a, one sample input image is shown where highlight and shadow are present. Fig. 7b shows the albedo $\rho$ image produced by our EM method, and Fig. 7c depicts the color-coded normal map produced by our EM method, where $(R, G, B) = \left(\frac{x^2}{2}, \frac{x^1}{2}, z\right)$ and $n = (x, y, z)^T$. Fig. 7d shows the normal map produced by our EM method, which is compared with the final result produced by [32] in Fig. 7e, where all surface details are smoothed out. Although the teapot in Fig. 7e demonstrates a visually smoother appearance, all the fine details described in Fig. 4 are lost due to the MRF refinement process. On the other hand, the normals produced by our EM method preserve all important fine details of the *Teapot* illustrated and revealed in the close-up views of Fig. 4.

We applied our method to more data sets and the results are also shown in Fig. 7. In these complex cases, our method works very well in estimating the surface albedos and surface normals. There is hardly any shading left in the albedo image of *Toy Car*. For *Rope*, only some small black

\[
R_i = \frac{1}{\sum_j v_{ij}} \left( v_{li}(h_i + p_i) + v_{ri}(h_R - p_R) + v_{ti}(h_T - q_i) + v_{bi}(h_B - q_B) \right). \tag{28}
\]

Fig. 8. Slope: (a) A partially integrable surface (ground truth). (b) The corresponding needle map. (c) Color-coded input normal map. (d) The weighted discontinuity map $\nu$ found by our EM algorithm. (e) Surface reconstructed by using integration. (f) Surface reconstructed by using [9]. (g) Surface reconstructed by using [19]. (h) Surface reconstructed by using our method. The root-mean-square errors of (e), (f), (g), and (h) are 50.0095, 37.6543, 48.0588, and 0, respectively.

2. For better visualization, the difference-image is remapped by reversing its color and then increasing the value of the red channel by 44.
spots are left in the albedo image because these spots were always under shadow due to the complex mesostructure. The surface normal maps also suffer fewer artifacts (e.g., the lower right part under the steering wheel of Toy Car) than the map obtained in [32]. For Face, the normal map shows that our method retains better the subtle geometric details, such as the pimple and other facial imperfections. The object demonstrated in Transparency exhibits little Lambertian property, which presents a lot of challenges because the toy is contained inside an open box, which casts a lot of shadows when the object is illuminated toward each side of the box. Note that the toy is wrapped inside a transparent plastic container. So, when it is illuminated from the top, a lot of highlight is produced. Despite these challenges, note the faithfulness of the normals and albedos recovered by our EM-based method.

6.2 Discontinuity Map and Height Field
The height field generation experiments were performed on a Pentium D 2.8 Hz machine with 1 GB memory.

Slope. We show in Fig. 8 a synthetic example Slope. Note that this example is different from and more difficult than the one shown in [19], where the surface transition is smooth along the slab boundary, whereas ours consists of $C^0$ and/or $C^1$ discontinuities (Figs. 8b and 8c). Figs. 8e, 8f, and 8g are the surfaces produced by integration, [9] and [19], respectively. The Matlab implementation of [9] and [19] was obtained from the author of [19] and is gratefully acknowledged. All of these methods fail to produce a surface faithful to the ground truth shown in Fig. 8a because it is incorrect to globally enforce the integrability constraint. Specifically, the horizontal plane should not contribute to the inclined plane in the surface integration process. Using our method, a weighted discontinuity map $v$ is estimated (Fig. 8d), which is used to weigh the contribution of...
neighbors. Our result, shown in Fig. 8h, is faithful to the ground truth.

**Three spheres.** Fig. 9 shows another example. Again, the methods in [9] and [19] produce distorted results. Note that a binary discontinuity map is inadequate (Fig. 9e): The small spheres observed have $C^0$ continuity and $C^1$ discontinuity (that is, sharp changes in surface gradients) along the loci of contact with the large sphere, except along the $C^0$ discontinuity (that is, the occlusion boundary), where the small spheres tangentially occlude the large one.

By setting all $v_{ij} = 1$ and applying only our height field estimation method, which is an analogy of [11], we still produce a decent surface shown in Fig. 9f. Although distortion can be observed around the small spheres and the contact between the plane and the large sphere, the salient features are still retained. Fig. 9g shows the result produced by our EM method, where less distortion is observed. Our result is comparable to the ground truth (Fig. 9h), showing the necessity of our weighted discontinuity map. Overall, the weighted discontinuity map is capable of dealing with this complex situation, restricting the influence of the largest sphere’s contribution to the small spheres along the loci of contact to recover the faithful depths for the small spheres. Figs. 9h and 9i, respectively, show the side views of our result and the ground truth for comparison.

**Noise robustness.** From left to right, the first row in Fig. 10 shows the input, respectively, corrupted with 0, 5, 10, 15, and 20 percent Gaussian noise. The second row shows the corresponding discontinuity map $v$ estimated. The third row shows the reconstructed height fields. Note that the distortion of the small spheres becomes apparent starting from the 10 percent case. However, even in the presence of 20 percent Gaussian noise, salient features are still maintained despite the observable distortion. Observe that we do not smooth out but preserve the “noise” on the reconstructed surfaces in the 5 percent case, indicating that our approach can better preserve fine details or subtle geometry found in real-life objects.

3. We take the noiseless color-coded normal map and use PhotoShop to add noise to it before converting back the noisy color-coded images, as shown in Fig. 10, to 3D normals.
Real examples. Fig. 11 shows the discontinuity maps and the height fields on some real examples, showing that our method can produce more reasonable results. The results produced by [19], [9], on the other hand, exhibit noticeable artifacts: The surfaces of Toy Car are significantly distorted; the mouth protrudes unnaturally in the Face surface results. In particular, note that the resulting surface of Transparency suffers a curved distortion. This resulted from the camera distortion (perspective and radial distortion), which makes the recovered surface normals tend to point toward the center-of-projection. Also, occluded and disconnected surfaces cannot be captured by a single view, which are manifested into surface distortion, a fundamental problem in the normal-to-surface conversion process.

Fig. 12. Teapot: (a) Color-coded normal map. (b) The weighted discontinuity map computed by our EM algorithm. (c) The surface reconstructed by using [19]. (d) The surface reconstructed by using our method. (e) The real teapot captured with a similar view point as (c) and (d). (f) The zoom-in view of the handle of (c). (g) The zoom-in view of the handle of (d). (h) The zoom-in view of the real teapot handle. (i) The zoom-in view of the lid of (c). (j) The zoom-in view of the lid of (d). (k) The zoom-in view of the real teapot.

Fig. 13. Male Torso. The top and bottom, respectively, show the front and back torsos. From left to right: One of the input images, the albedo estimated by EM, the color-coded normal map and the Lambertian-shaded normal map produced by our EM method, assuming the light direction $L = \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right)^T$, the discontinuity map estimated by EM, and the frontal-view surface.
Although our EM algorithm soft-labels orientation and depth discontinuities, the surface reconstruction result is still a singly connected surface, which cannot handle occlusion and depth discontinuities.

We compare our result with the result produced using [19] (Figs. 12c and 12d). Our result exhibits significantly less distortion. Figs. 12f, 12g, 12h, 12i, 12j, and 12k show the zoom-in views of selected regions. Compared with the real object, our result is more reasonable. Note that the lid in Fig. 12i is clearly distorted, where the teapot lid appears stretched producing an asymmetric appearance. On the other hand, our method faithfully maintains the shape and preserves the subtle geometry of the teapot without significant distortion. In Figs. 13 and 14, we show the reconstruction results of albedo, normal map, discontinuity map, and surface of two real statuettes. Note the details recovered in the results.

7 Conclusion
This paper presents a complete pipeline based on EM to address the photometric stereo reconstruction problem. From a dense and noisy set of photometric images captured using a simple setup, our algorithm produces normals, albedos, and surfaces faithful to those of the captured real object. The key ideas lie in 1) alternatingly optimizing the selection of Lambertian observations, the estimation of surface normals, and albedos to reconstruct accurate normals from dense images and 2) alternately optimizing surface integrabilities and discontinuities to reconstruct a visible surface from the optimized normals. Our detailed derivation produces very simple updating rules for the EM algorithms, allowing a stable and practical system be built that has no free parameters to set. We demonstrated very good results despite the presence of non-Lambertian observations and complex reflectance, which are both problematic to previous methods in calibrated photometric stereo. It will be a fruitful attempt to explore the issues in constructing a real-time photometric stereo system based on the algorithms of this paper, which will be one focus of our future work. Multiview photometric stereo, which can be used to recover high-quality 3D model, is another research direction for future pursuit.

Acknowledgments
The authors would like to thank all of the anonymous reviewers for their thoughtful comments and constructive suggestions. This research is supported by the Hong Kong Research Grant Council (grant number 620309).

References