Dense Photometric Stereo by Expectation Maximization*

Tai-Pang Wu and Chi-Keung Tang

Vision and Graphics Group,
The Hong Kong University of Science and Technology,
Clear Water Bay, Hong Kong

Abstract. We formulate a robust method using Expectation Maximization (EM) to address the problem of dense photometric stereo. Previous approaches using Markov Random Fields (MRF) utilized a dense set of noisy photometric images for estimating an initial normal to encode the matching cost at each pixel, followed by normal refinement by considering the neighborhood of the pixel. In this paper, we argue that they had not fully utilized the inherent data redundancy in the dense set and that its full exploitation leads to considerable improvement. Using the same noisy and dense input, this paper contributes in learning relevant observations, recovering accurate normals and very good surface albedos, and inferring optimal parameters in an unifying EM framework that converges to an optimal solution and has no free user-supplied parameter to set. Experiments show that our EM approach for dense photometric stereo outperforms the previous approaches using the same input.

1 Introduction

Woodham [1] first introduced photometric stereo for Lambertian surfaces, in which three images are used to solve the reflectance equation for recovering surface gradients and albedos of a Lambertian surface. Since [1], extensive research on more robust techniques for photometric stereo have been reported:

More than three images. Four images were used in [2] and [3] so that inconsistent observation due to shadows or highlight can be discarded by majority vote. A larger number of images (about 20) were used in [4] where two algorithms were investigated. More recently, [5] used structure from motion and photometric stereo in an iterative framework.

Model-based approaches. In [6], an m-lobed reflective map was derived by considering diffuse and non-Lambertian surfaces. This was extended in [7] in which nonlinear regression was applied to a larger number of input images. The Torrance-Sparrow model was used in [3]. In [8], a hybrid reflectance model was used to recover surface gradients and the parameters of the reflectance model.

* This work was supported by the Research Grant Council of Hong Kong Special Administrative Region, China, under grant # 62005.

A. Leonardis, H. Bischof, and A. Pinz (Eds.): ECCV 2006, Part IV, LNCS 3954, pp. 159–172, 2006.

Reference objects. The use of a reference object was first introduced in [9]. In [10], surface orientations and reflectance properties are computed by using a reference object.

Despite that significant advancement has been made in photometric stereo by previous approaches, they still suffer from one or more of the following limitations:

- light directions must be very accurate. The use of uncalibrated lights require additional constraints [11].
- accurate normals and albedos cannot be recovered in the presence of highlight and cast shadows, and severe violations to the Lambertian assumption.
- in certain model-based approaches, the problem formulation is very complex, making them susceptible to numerical instability.

Recently, two Markov Random Field (MRF) inference algorithms [12, 13] were developed independently to recover normals by dense photometric stereo using a dense set of noisy photometric images conveniently captured by a simple setup. These two methods were based on similar MRF formulation but different distribution models, and made use of the neighborhood information to improve the results. For high precision normal reconstruction, the graph-cut algorithm [13] converges in a few iterations. The tensorial message passing was proposed in [12] for efficient belief propagation. In both cases, estimated normal maps are very good (certain subtle geometry can be reasonably reconstructed) despite the presence of highlight, shadows and complex geometry. Albeit this, several issues remain unaddressed:

- Albedo is not recovered in [12, 13].
- The data redundancy inherent in the dense set has not been fully utilized. Specifically, linear plane fitting was used to estimate an initial normal at each pixel based on the assumption that sufficient linear Lambertian observations are present. However, each observation, regardless of Lambertian or otherwise, is equally weighted during the plane fitting process.
- The MRF is introduced in [12, 13] to improve the results. However, the introduction implies the surface smoothness assumption. Despite the use of a discontinuity-preserving metric, [12, 13] apply the MRF refinement globally as in other MRF methods. Very fine details such as subtle texture bumps and surface imperfections will inevitably be lost after the process.
- As with other MRF processes, a user-supplied parameter is required to control the influence of neighborhood. The optimal parameter is different for different scenes and has to be determined empirically but not automatically.

In this paper, we propose a unifying framework based on the Expectation Maximization (EM) algorithm to address all the above seven issues. We shall show that considerable improvement are made by our EM approach, using the same noisy dense set as input. The organization of the paper is as follows. Section 2 reviews dense photometric stereo and describes the above issues in detail in order to motivate our work. Section 3 describes our unified EM framework. Finally, results are presented in Section 4 and we conclude our paper in Section 5.

2 Review of Dense Photometric Stereo

Given a dense set of images captured at a fixed viewpoint with their corresponding light directions, the goal of dense photometric stereo is to find the optimal normal N_s and albedo ρ_s at each pixel s. In [13], a simple capture device was proposed for obtaining a dense but noisy set of photometric images. By utilizing the redundancy inherent in the captured data, a dense matching cost was derived and used as the local evidence at each observation node in the MRF network. The capture process is simple [13] compared with other previous approaches whereas the methods in [12, 13] produced some of most accurate normal reconstruction to date despite the presence of severe shadows, highlight, transparencies, complex geometry, and inaccurate estimation in light directions.

2.1 Data Acquisition

Shown in Fig. 1(a) is the simple capture system consisting of a digital video camera (DV), a handheld spotlight and a mirror sphere which is used to give the light direction. The location of the brightest spot on the mirror sphere indicates the light direction which can be calculated easily [13]. Note however that the set of estimated light directions is scattered and very noisy, as shown in Fig. 1(c). Uniform resampling on a light direction sphere, based on icosahedron subdivision, was performed in [13], which is also adopted in this work. The inaccurate light directions and the contaminated photometric images both make the reconstruction problem very challenging.

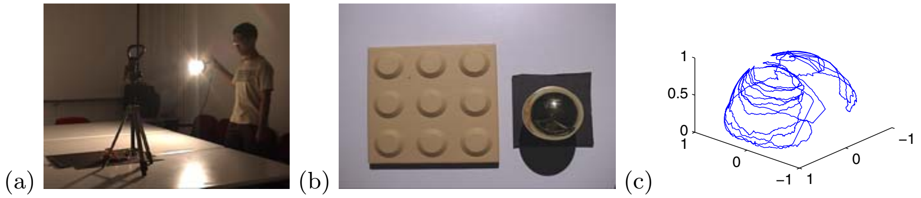

Fig. 1. (a) Data capture. (b) Typical captured image. (c) A typical trajectory of the estimated light directions shows that they are scattered and very noisy.

2.2 The MRF Formulation

Given a set of photometric images with the corresponding estimated light directions, the surface normals are estimated by maximizing the following posterior probability [12, 13]:

$$P(X|Y) \propto \prod_{s} \varphi_s(x_s, y_s) \prod_{s} \prod_{t \in \mathcal{N}(s)} \varphi_{st}(x_s, x_t) \qquad (1)$$

where $X = \{x_s\}$, $Y = \{y_s\}$, x_s is the hidden variable (i.e. the normal to be estimated) at pixel location s, y_s is the observed normal at s, $\mathcal{N}(s)$ is a set of first order neighbors of s, and $\varphi_s(x_s, y_s)$ is the *local evidence* at the observation node and $\varphi_{st}(x_s, x_t)$ is the *compatibility function.* To maximize (1), tensorial belief propagation was used in [12] while graph-cut was used in [13].

2.3 Issues in Deriving Local Evidences

In [12, 13] the local evidence and initial normals are derived by least-square plane fitting, assuming that sufficient Lambertian observations are present and that non-Lambertian observations are noises. The reflectance at each pixel can then be described by $\rho(\mathbf{N}_s \cdot \mathbf{L})$, where ρ is the surface albedo, $\mathbf{N}_s$ is the initial normal and $\mathbf{L}$ is the light direction at the pixel s. Let T be the total number of sampled images. To eliminate ρ, we divide $T-1$ sampled images by a chosen image called *denominator image* to obtain $T-1$ *ratio images.* Let I_d be the denominator image. Each pixel in a ratio image is therefore expressed by $\frac{I_t}{I_d} = \frac{\mathbf{N}_s \cdot \mathbf{L}_t}{\mathbf{N}_s \cdot \mathbf{L}_d}$. By using no less than three ratio images, we produce a local estimation of the normal at each pixel:

$$A_t x + B_t y + C_t z = 0 \tag{2}$$

where $A_t = I_t l_{d,x} - I_d l_{t,x}$, $B_t = I_t l_{d,y} - I_d l_{t,y}$, $C_t = I_t l_{d,z} - I_d l_{t,z}$, $\mathbf{L}_t = (l_{t,x}, l_{t,y}, l_{t,z})^T$ is the light direction at time $t = 1 \cdots T$, $\mathbf{N}_s = (x, y, z)^T$ is the normal to be estimated. Note that an ideal denominator image is one that satisfies the Lambertian model and is minimally affected by shadows and specular highlight, which is difficult if not impossible to obtain. On the other hand, the use of least-square plane fitting to estimate $\mathbf{N}_s$ has several problems:

- The albedo is canceled out to produce ratio images.
- Least-square plane fitting is incapable of rejecting non-Lambertian observations. Outliers significantly affect the result of the fitting.
- If the denominator contains a non-Lambertian observation, the whole set of ratio images becomes garbage thus leading to unpredictable results.

Different alternatives of selecting the denominator image from the dense set have been proposed. Fig. 2(a) shows the normal map produced by [13]. Problems can be observed on the "ground" in the bottom part of the teapot image because only a *single* image is used which is chosen using simple criteria. Fig. 2(b) shows another initial normal map produced by [12]. The result is very noisy because in [12] *different* images were used as denominators for different pixels. Severe orientation jittering is resulted in the estimated normal map due to the non-Lambertian properties and the quantization errors of the intensities of the denominator image.

So the first question we ask is: *can we identify or learn the relevant Lambertian observations automatically?*

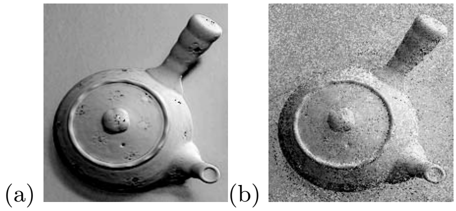

Fig. 2. Local evidence (initial normals) produced by (a) [13] and (b) [12]

2.4 Issues in Defining Compatibility Functions

In [12, 13] the noises due to non-Lambertian observations and inaccurate estimation in light directions that cannot be handled by plane fitting are addressed by the MRF refinement process which assumes that the underlying surface is locally smooth. Although discontinuity-preserving functions are used, the smoothing effect is applied globally because we have no prior knowledge which regions should be smoothed. If the variation in global and local surface orientation do not match, over-smoothing will occur. Therefore, fine details such as surface imperfections and texture bumps are inevitably lost. Thus, in [12, 13] and other MRF algorithms, a free parameter should be supplied by the user to control the degree of smoothness. The parameters are empirically obtained and varies with different scenes.

So, the second question we ask is: *how can we obtain the set of optimal parameters automatically?*

3 Normal and Albedo Estimation by Expectation Maximization

In this paper, a unified EM algorithm is proposed which identifies relevant Lambertian observations automatically by fully exploiting the data redundancy inherent in the dense and noisy data. Our results show significant improvement without any MRF smoothing refinement and thus the setting of MRF parameters is no longer an issue. In fact, by using our EM algorithm, *all* parameters can be optimized alternately within the same framework, making the robust method free of any user-supplied parameters.

In this section, we formulate our EM algorithm to estimate the surface albedos and normals from a set of dense and noisy measurement captured as described in the previous section. In [12, 13], the simple least-square plane fitting is used for initial normal estimation. No special handling is performed for unreliable data or outliers generated by non-Lambertian phenomena such as specular highlight and shadows. In real cases, however, these observations occupy a significant proportion in the captured data due to the restrictive Lambertian model and the diversity of surface geometry and material.

Suppose that the measurement error for each observation is known. We could perform *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. In this paper we propose a data-driven approach to estimate the weight of each observation by utilizing useful information inherent in the dense set although it consists of scattered and noisy data.

3.1 Overview

While the albedo is problematic and canceled out in [12, 13], in this paper 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 the albedo ρ_i is known, given the observed intensity I_{it} at time t and the corresponding light direction $\mathbf{L}_{it}$, we model the probability of the intensity I_{it} generated by the Lambertian model without shadow and specular highlight to be inversely proportional to:

$$||I_{it} - \rho_i \mathbf{N}_i \cdot \mathbf{L}_{it}|| \tag{3}$$

where $\mathbf{N}_i$ is the normal at pixel i. Thus, if the albedo is known, more information concerning the observations can be extracted.

However, albedo derivation alone is a difficult problem. In this paper, we demonstrate how the albedo and surface normal can be estimated *simultaneously* using an EM framework to obtain accurate results.

While we argue that plane fitting without a proper contribution weight for each observation is not a good solution, some useful lesson can still be learnt from [12, 13]. Suppose that each image is a candidate of the denominator image. If we have T different observations for a pixel location, in total, we can produce T different planes by using all images successively as the denominator. For the denominators consisting of non-Lambertian observations, the orientations of the fitted planes are arbitrary because the denominator intensity interacts with all other intensity samples when ratio images are derived during plane fitting and thus the whole data set is contaminated. For the other denominators whose observations are explained by the Lambertian model, however, the orientations of the produced planes should cluster themselves together. Despite that such estimated planes are not error-free because of the presence of outliers, the cluster limits the solution space for the optimal surface orientation at the pixel.

3.2 The Objective Function

The main reason of using the EM approach is that the above-mentioned cues are not given but are inherent within the data itself. Alternating optimization approaches such as EM allow for the simultaneous estimation of the cues and the solution. In this section, we define our objective function which forms the basis of our EM algorithm.

Without confusion, in the rest of this section, the index of pixel location i will be dropped to simplify the notation, since the algorithm is applied individually at each pixel location.

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

To encode the clustering of $\{\mathbf{n}_t\}$, a 3×3 covariance matrix $\mathbf{K}$ that stores the second-order moment collection is used since it represents the orientation distribution. The optimal normal is the direction that gives the largest variance in $\mathbf{K}$.

[1] Note that we only use the highest 50% intensities as numerators to perform plane fitting because dark pixels tend to be affected by shadows and the presence of a large number of outliers will affect the accuracy of the estimated normal. The aforementioned number of samples provides sufficient redundancy for robust estimation.

Our goal is to find the optimal albedo ρ (a 3-vector in RGB space) and the covariance matrix $\mathbf{K}$ given the pixel observations. In other words, we want to estimate the following:

$$\Theta^* = \arg\max_{\Theta} P(\mathbf{O}, \mathbf{S}|\Theta) \tag{4}$$

where $P(\mathbf{O}, \mathbf{S}|\Theta)$ is the complete-data likelihood we want to maximize, $\Theta = \{\mathbf{K}, \rho, \alpha, \sigma\}$ is a set of parameters to be estimated and $\mathbf{S} = \{s_t\}$ 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. α and σ are respectively the proportion of Lambertian observations and the standard deviation of Eqn. 3, which are the parameters that help us to find the solution and will be described.

Our EM algorithm estimates Eqn. 4 by finding the expected value of the complete-data log-likelihood $\log P(\mathbf{O}, \mathbf{S}|\Theta)$ w.r.t. $\mathbf{S}$ given the observation $\mathbf{O}$ and the current estimated parameters:

$$Q(\Theta, \Theta') = \sum_{\mathbf{S} \in \varphi} \log P(\mathbf{O}, \mathbf{S}|\Theta) P(\mathbf{S}|\mathbf{O}, \Theta') d\mathbf{S} \tag{5}$$

where Θ' are current parameters and φ is a space containing all $\mathbf{S}$ of size T.

3.3 Expectation Estimation

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 defined by Eqn. 5 by proceeding to the next iteration given the current parameters.

If s_t is known, the observation o_t that is generated by the Lambertian model minimizes Eqn. 3 and $\mathbf{n}_t^T \mathbf{K}^{-1} \mathbf{n}_t$. Suppose that the noise distribution of Eqn. 3 and the jittering distribution of $\mathbf{n}_t$ are Gaussian distributions, and that the existence of non-Lambertian observations follow a uniform distribution. The observation probability of o_t is:

$$p(o_t|s_t, \Theta') \propto \begin{cases} \exp(-\frac{||I_t - \rho \mathbf{n}_t \cdot \mathbf{L}_t||^2}{2\sigma^2}) \exp(-\frac{1}{2}\mathbf{n}_t^T \mathbf{K}^{-1} \mathbf{n}_t), & \text{if } s_t = 1; \\ \frac{1}{C}, & \text{if } s_t = 0. \end{cases} \tag{6}$$

Base on the uniform distribution assumption, the choice of C should be $max\{I_t - \rho \mathbf{n}_t \cdot \mathbf{L}_t\}$. However, in real case, the assumption can be violated seriously. To lower the chance of wrong classification, we choose $C = C_m = mean\{I_t - \rho \mathbf{n}_t \cdot \mathbf{L}_t\}$ because smaller C trends to classify more observations to $s_t = 0$. This lowers the probability of the non-Lambertian samples in obtaining wrong labels while we still have sufficient redundancy for estimation robustness. To calculate C, we choose ρ to be the color has 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 α be the proportion of the observation generated by the Lambertian model. Then we have a mixture probability of the observations:

$$p(s_t = 1) = \alpha \tag{7}$$

So, given Θ' only, we have

$$p(o_t|\Theta') \propto \alpha \ \exp(-\frac{||I_t - \rho \mathbf{n}_t \cdot \mathbf{L}_t||^2}{2\sigma^2}) \exp(-\frac{1}{2}\mathbf{n}_t^T \mathbf{K}^{-1} \mathbf{n}_t) + \frac{1-\alpha}{C} \tag{8}$$

Let w_t be the probability of o_t being generated by the Lambertian model. Then:

$$\begin{aligned} w_t = p(s_t = 1|o_t, \Theta') &= \frac{p(o_t, s_t = 1|\Theta')}{p(o_t|\Theta')} \\ &= \frac{\alpha \ \exp(-\frac{||I_t - \rho \mathbf{n}_t \cdot \mathbf{L}_t||^2}{2\sigma^2}) \exp(-\frac{1}{2}\mathbf{n}_t^T \mathbf{K}^{-1} \mathbf{n}_t)}{\alpha \ \exp(-\frac{||I_t - \rho \mathbf{n}_t \cdot \mathbf{L}_t||^2}{2\sigma^2}) \exp(-\frac{1}{2}\mathbf{n}_t^T \mathbf{K}^{-1} \mathbf{n}_t) + \frac{1-\alpha}{C}} \end{aligned} \tag{9}$$

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

3.4 Maximization

In this section, we maximize the likelihood (Eqn. 4) 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 function (Eqn. 5) is:

$$\begin{aligned} Q(\Theta, \Theta') &= \sum_t \log p(o_t, s_t = 1|\Theta) w_t + \sum_t \log p(o_t, s_t = 0|\Theta)(1 - w_t) \\ &= \sum_t \log(\alpha \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{||I_t - \rho \mathbf{n}_t \cdot \mathbf{L}_t||^2}{2\sigma^2})) w_t \\ &+ \sum_t \log(\frac{1}{|\mathbf{K}|^{\frac{1}{2}} (2\pi)^{\frac{3}{2}}} \exp(-\frac{1}{2}\mathbf{n}_t^T \mathbf{K}^{-1} \mathbf{n}_t)) w_t \\ &+ \sum_t \log(\frac{1-\alpha}{C})(1 - w_t) \end{aligned} \tag{10}$$

To maximize (10), we set the first derivative of Q w.r.t. α, σ, ρ and $\mathbf{K}$ respectively equals to zero and obtain the following:

$$\begin{aligned} \alpha &= \frac{1}{T} \sum_t w_t \\ \sigma &= \frac{\sum_t ||I_t - \rho \mathbf{n}_t \cdot \mathbf{L}_t||^2 w_t}{\sum_t w_t} \\ \rho &= \frac{1}{\sum_t (\mathbf{n}_t \cdot \mathbf{L}_t)^2 w_t} \sum_t I_t (\mathbf{n}_t \cdot \mathbf{L}_t) w_t \\ \mathbf{K} &= \frac{1}{\sum_t w_t} \sum_t \mathbf{n}_t \mathbf{n}_t^T w_t \end{aligned} \tag{11}$$

which constitutes the parameter updating rule for Θ and thus the M-Step of our EM algorithm. The E-Step and M-Step are executed alternately until the process converges. The convergence of EM was well established [14].

Upon convergence, we apply eigen-decomposition on $\mathbf{K}$ to obtain the optimal normal direction. The eigenvector corresponding to the largest eigenvalue gives the normal direction.

In addition, using our method, we produce not only surface normals but also surface albedo ρ and the weights w_t indicating the degree an observation o_t is consistent with the Lambertian model. Such inferred information is very useful in parameter estimation for fitting analytic reflectance models to real and noisy observations.

4 Experimental Results

In this section, we first demonstrate the considerable improvement by comparing our method with [13] using the same input data. The synthetic case we use is *Three Spheres* and the real examples are *Teapot*, *Rope* and *Toy Car*. After the comparison, we apply our method to reconstruct albedos and normals on selected complex objects to examine the robustness and efficacy of our method. The running time of all the examples are tabulated in Table 1.

Three Spheres. Fig. 3(a)–(b) show two input synthetic images of *Three Spheres*. The depicted object is generated by the Phong illumination model. Fig 3(e) and (f) shows respectively the normal map produced by [13] and by our EM method. Note that they are rendered using the Lambertian model ($\mathbf{N}\cdot\mathbf{L}$) for clarity of display. The ground truth is shown in Fig. 3(c). Our estimated albedo is shown in Fig. 3(d). Qualitatively, the appearance of (c), (e) and (f) are very similar. On the other hand, the image difference show the clear improvement of our method in terms of accuracy. Fig. 3(g) is the image difference between (c) and (e), while Fig. 3(h) is the image difference between (c) and (f). Notice the presence of three halos in Fig. 3(g) which are brighter than those observed in Fig. 3(h), which is nearly totally black. 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 [13] is 4.041 degree while the error of our EM result is only 1.5065 degree.

Teapot. Our method shows very significant improvement in the presence of a large amount of noises in the representative case of *Teapot*, which is one of the most difficult examples in [13] where the geometry and texture are very complex.

Table 1. Summary of running times. The experiments were run on a shared CPU server with 4 Opteron(TM) 844 CPU at 1.8GHz with 16GB Memory.

Data set	*Three Spheres*	*Teapot*	*Toy Car*	*Rope*	*Face*	*Hair*
Number of images	305	282	287	265	195	189
Image Size	256x256	188x202	181x184	171x144	216x225	224x298
Running Time	3m04s	4m24s	4m37s	4m09s	4m45s	3m25s

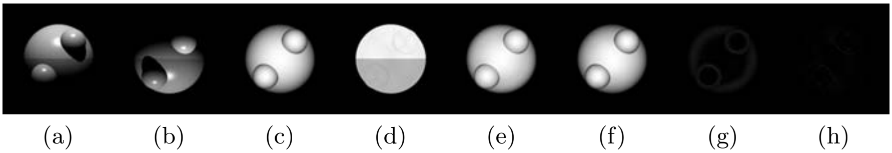

Fig. 3. *Three Spheres*: (a)–(b) Two typical noisy input images. (c) The ground truth normal map. (d) The albedo ρ produced by our EM method. (e) The normal map produced by [13]. (f) The normal map produced by our EM method. (g) The image difference between (c) and (e). (h) the image difference between (c) and (f). Note that (c), (e) and (f) are rendered using the pure Lambertian model ($\mathbf{N}{\cdot}\mathbf{L}$) with $\mathbf{L} = (0, 0, 1)^T$.

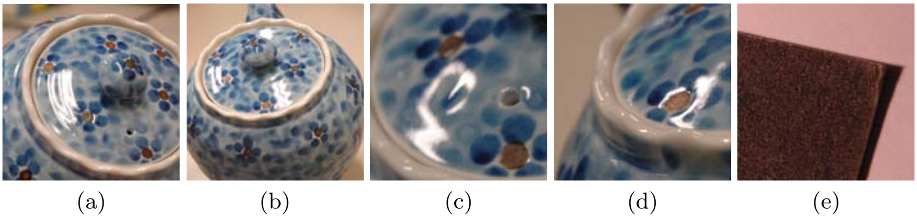

Fig. 4. Detail of the *Teapot*: (a)–(b) The specular reflection depicts the concentric ripple-shaped structures on the lid. (c) The specular reflection depicts a smooth but shallow *dent* near the hole of the lid. (d) A small *bump* at the center of a deep-colored flower pattern. (e) A black cardboard with a lot of *surface imperfections*, which is the plane where the teapot is placed for image capturing.

To better illustrate how our result has been improved, let us study in detail the geometry of the *Teapot* using Fig. 4. The selected close-up views of the teapot reveal fine surface details and subtle geometry.

The complete set of the result shown in Fig. 5. Fig. 5(a) and (b) show two sample input images which are contaminated by highlight and shadows. To show the overall smoothness, Fig. 5(c) depicts the color coded normal map produced by our EM method where $(R, G, B) = (\frac{x+1}{2}, \frac{y+1}{2}, z)$ and $\mathbf{N} = (x, y, z)^T$. Fig. 5(d) is the albedo ρ image produced by our EM method. Fig. 5(e) shows the local evidence of [13] which consists of the initial normals produced by the plane fitting method reviewed in Section 2 or described in [13]. Fig. 5(g) is the final result produced by [13] where all surface details are smoothed out. For clarity of display, Fig. 5(f) and (h) show the same normal map produced by our EM method, which are rendered using the Lambertian model ($\mathbf{N} \cdot \mathbf{L}$) illuminated at two different light directions.

It is evident that, although Fig. 5(g) demonstrates a visually smoother appearance, all the fine details described in Fig. 4 are lost due to the MRF refinement process. However, Fig. 5(e) show that if MRF process is not applied in [13], due to the complexity of the texture and the geometry, the surface normals produced are unsatisfactory and severe artifacts can be observed. On the other hand, our method preserves all important fine details of the *Teapot* illustrated and revealed in the close-up views of Fig. 4.

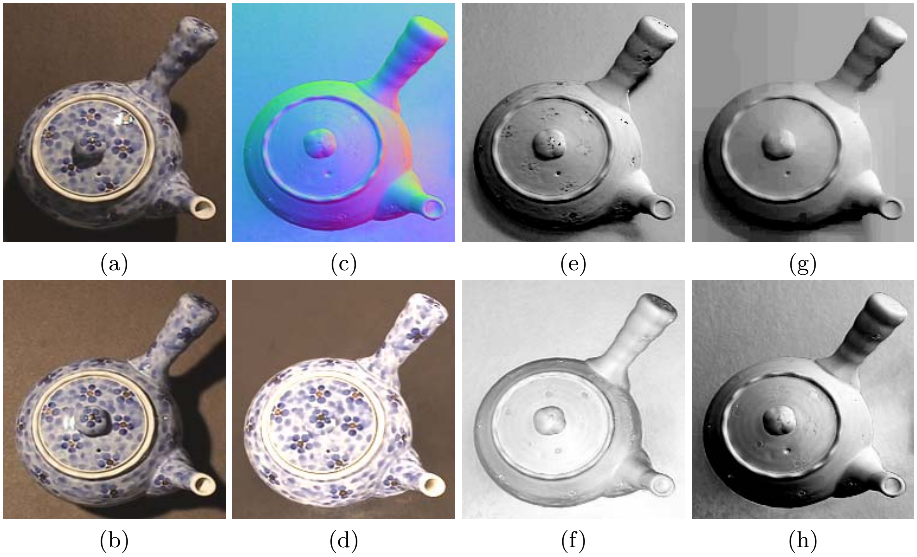

Fig. 5. *Teapot*: (a)–(b) Two captured images. (c) The color coded normal map produced by our EM method. (d) The albedo ρ estimated by our EM method. (e) The local evidence of [13]. (g) The final normal map in [13]. (f) and (h) are the same normal map produced by our EM method. Note that the normal maps in (e)– (h) are rendered using the pure Lambertian model ($\mathbf{N} \cdot \mathbf{L}$) where the light directions in (e), (g) and (h) are respectively $\mathbf{L} = (\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})^T$ and in (f) is $\mathbf{L} = (0, 0, 1)^T$. Please see the electronic version for higher resolution display.

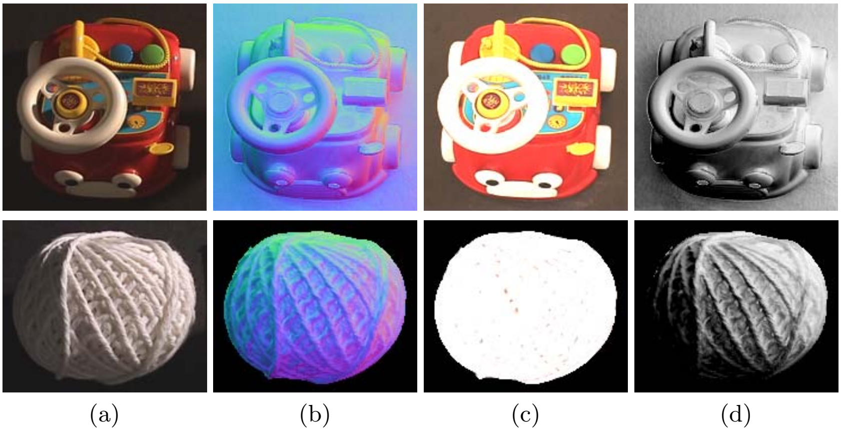

Fig. 6. Toy Car (first row) / Rope (second row) : (a) One of the input images. (b) The color coded normal map produced by our EM method. (c) The albedo ρ estimated by our EM method. (e) The normal map produced by our EM method rendered by the pure Lambertian model ($\mathbf{N} \cdot \mathbf{L}$) where the light direction $\mathbf{L} = (\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})^T$.

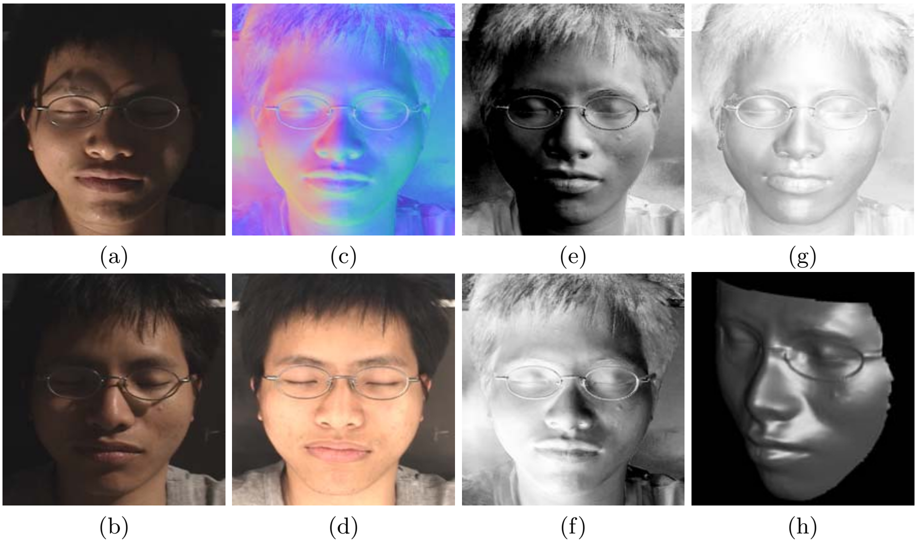

Fig. 7. *Face*: (a)–(b) Two captured images. (c) The color coded normal map produced by our EM method. (d) The albedo ρ estimated by our EM method. (e)–(g) The normal map produced by our EM method rendered using the Lambertian model $(\mathbf{N} \cdot \mathbf{L})$ where the light direction in (e), (f) and (g) are $\mathbf{L} = (\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})^T$, $\mathbf{L} = (-\frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})^T$ and $\mathbf{L} = (0, 0, 1)^T$ respectively. (h) The reconstructed surface.

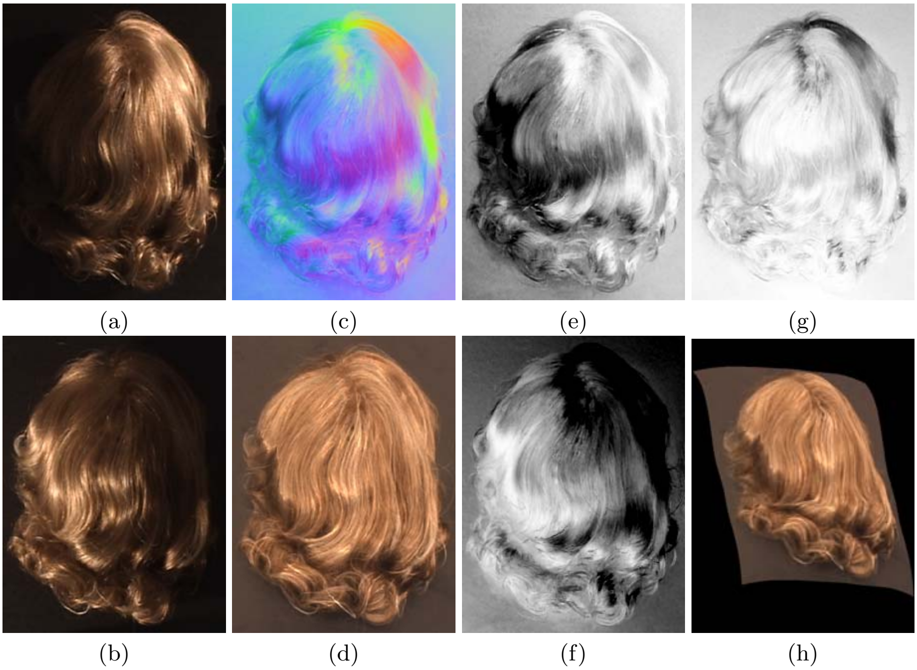

Fig. 8. *Hair*: Please see the caption in Fig. 7

Toy Car and Rope. We applied our method to two existing data sets and the results are shown in Fig. 6. Our method works very well in estimating the surface albedos and surface normals. Hardly any shading is left in the albedo image of *Toy Car*. For *Rope*, only some small spots of shadow artifact are left in the albedo image because these regions were always under shadow due to the complex mesostructure. Besides, the surface normal maps obtained are more accurate than the maps obtained in [13].

Face and Hair. Human face and hair reconstruction are receiving more attention in the area of computer vision and computer graphics. Both human features consist of complex geometry and fine details and are non-Lambertian.

Fig. 7 shows our result on *Face*. The normal map (Fig. 7(e)–(g)) shows that our method retains the subtle geometry such as the pimple and other facial imperfections. Fig. 7(d) shows the estimated albedo image. Fig. 7(h) depicts the reconstructed surface by [15] using our normal map as input.

Fig. 8 shows our result on *Hair*. The normal map (Fig. 8(e)–(g)) shows that our method preserves the curvilinearity and the meso-structural details of the hair. Observe that some structure information are left in the albedo image (Fig. 8(d)). There are two reasons. First, although the sampled light directions are very dense, some pixels are always occluded and thus under shadows due to the complexity of the hair geometry. Besides, the Lambertian model is not sufficient to describe human hairs and so Lambertian samples are rare even dense measurement is available. These two problems make the estimation process extremely challenging; yet our method still produces very good result in normal estimation.

5 Conclusion

In this paper we propose a robust method for dense photometric stereo reconstruction using the Expectation Maximization (EM). By exploiting useful information inherent in the dense and noisy set of photometric images, this paper contributes in identifying relevant observations, recovering very good normals and albedos, and estimating optimal parameters in an automatic EM framework that has no free user-supplied parameter to set. The convergence of the EM method has been well established. Very good results have been obtained, showing that our EM approach is robust in the presence of severe shadows, highlight, complex and subtle geometry, and inaccurate light directions. Our future work focuses on the use of adaptive MRF refinement to further improve the accuracy and applicability of our EM technique for photometric stereo.

References

1. Woodham, R.: Photometric method for determining surface orientation from multiple images. OptEng **19**(1) (1980) 139–144
2. Coleman, Jr., E., Jain, R.: Obtaining 3-dimensional shape of textured and specular surfaces using four-source photometry. CGIP **18**(4) (1982) 309–328

3. Solomon, F., Ikeuchi, K.: Extracting the shape and roughness of specular lobe objects using four light photometric stereo. PAMI **18**(4) (1996) 449–454
4. Lee, K., Kuo, C.: Shape reconstruction from photometric stereo. In: CVPR92. (1992) 479–484
5. Lim, J., Ho, J., Yang, M., Kriegman, D.: Passive photometric stereo from motion. In: ICCV05. (2005)
6. Tagare, H., deFigueiredo, R.: A theory of photometric stereo for a class of diffuse non-lambertian surfaces. PAMI **13**(2) (1991) 133–152
7. Kay, G., Caelly, T.: Estimating the parameters of an illumination model using photometric stereo. GMIP **57**(5) (1995) 365–388
8. Nayar, S., Ikeuchi, K., Kanade, T.: Determining shape and reflectance of hybrid surfaces by photometric sampling. IEEE Trans. on Robotics and Automation **6**(4) (1990) 418–431
9. Horn, B., Woodham, R., Silver, W.: Determining shape and reflectance using multiple images. In: MIT AI Memo. (1978)
10. Hertzmann, A., Seitz, S.: Shape and materials by example: a photometric stereo approach. In: CVPR03. (2003) I: 533–540
11. Drbohlav, O., Sara, R.: Unambiguous determination of shape from photometric stereo with unknown light sources. In: ICCV01. (2001)
12. Tang, K., Tang, C., Wong, T.: Dense photometric stereo using tensorial belief propagation. In: CVPR2005. Volume 1. (2005) 132–139
13. Wu, T., Tang, C.: Dense photometric stereo using a mirror sphere and graph cut. In: CVPR2005. Volume 1. (2005) 140–147
14. Bilmes, J.: A gentle tutorial on the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models. Technical Report ICSI-TR-97-021, ICSI (1997)
15. Kovesi, P.: Shapelets correlated with surface normals produce surfaces. In: ICCV05. (2005)