Dense Photometric Stereo Using a Mirror Sphere and Graph Cut *

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 present a surprisingly simple system that performs robust normal reconstruction by dense photometric stereo, in the presence of large shadows, highlight, transparencies, complex geometry, variable attenuation in light intensity and inaccurate light directions. Our system consists of a mirror sphere, a spotlight and a DV camera only. Using this, we infer a dense set of unbiased but noisy photometric data uniformly distributed on the light direction sphere. We use this dense set to derive a very robust matching cost for our MRF photometric stereo model, where the Maximum A Posteriori (MAP) solution is estimated. To aggregate support for candidate normals in the normal refinement process, we introduce a compatibility function that is translated into a discontinuity-preserving metric, thus speeding up the MAP estimation by energy minimization using graph cut. No reference object of similar material is used. We perform detailed comparison on our approach with conventional convex minimization. We show very good normals estimated from very noisy data on a wide range of difficult objects to show the robustness and usefulness of our method.

1 Introduction

Normal recovery by photometric stereo has regained interest because it may provide a space-efficient alternative to, for instance, bidirectional texture functions (BTF), face databases, and tensor textures for relighting and rendering applications, where in [21, 10, 22] the storage requirement is usually very high. While approaches in photometric stereo using two views with known albedos [23], three views [8], four views [6, 18, 3], more views [16], complex reflectance models [17, 19, 12, 18], lookup tables [23, 24], reference objects [9, 7], novel object representation [4] have been reported, photometric stereo is still considered to be a difficult problem in the presence of shadows, specular highlights, and objects with complex material and geometry.

Based on our theoretical MRF model for photometric stereo and a simple reflectance model, we propose a surprisingly simple and inexpensive system to derive the matching cost function from noisy observations. We use a simple model consisting of summing a specular and a Lambertian component. However, by employing dense and unbiased but noisy measurements, an effective normal recovery method is proposed and implemented, which is very robust to violations to many models assumed by previous work of photometric stereo. Very good results have been obtained. To illustrate, we show in Figure 1 some typical images obtained from our system. Large areas are contaminated by highlight and shadows. Nevertheless, we are able to recover very good surface normals. No reference object of

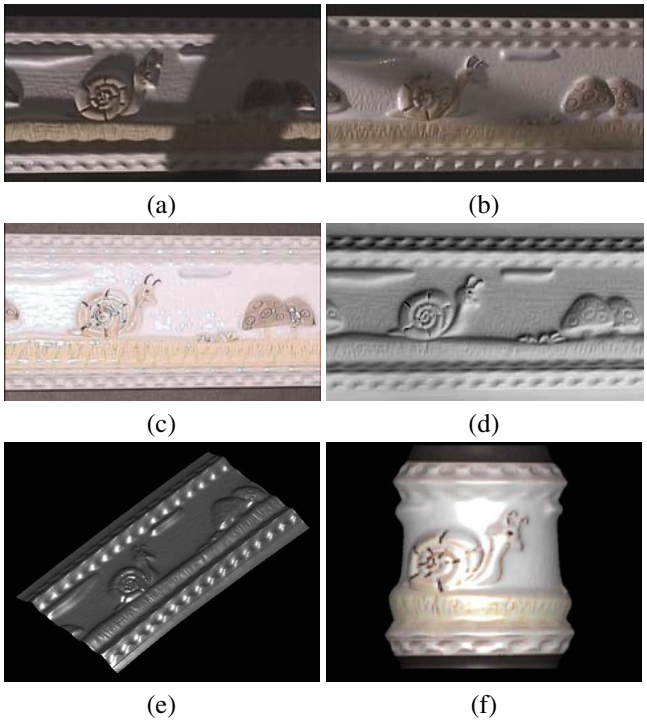


Figure 1: (a)-(c) three typical photometric images. (a) and (b) are significantly contaminated by shadows, and (c) is largely contaminated by highlight. (d) shows our recovered normals $\mathbf{N}$, which are displayed by $\mathbf{N} \cdot \mathbf{L}$ where $\mathbf{L} = (\frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})^T$. (e) Surface reconstructed from the recovered normals at a novel viewpoint. (f) 3D texture mapping on a cylinder, using the reconstructed surface and normals. Please see supplementary video [1].

*This research is supported by the Research Grant Council of Hong Kong Special Administration Region, China: AOE/E-01/99, HKUST6175/04E

Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)
1063-6919/05 $20.00 © 2005 **IEEE**

similar material with the target object is used. Our simple system consists of a distance spotlight to mimic a directional light source, a DV camera for image capturing, and a mirror sphere for light direction estimation.

Our MRF model for photometric stereo is translated into an energy function. Estimating the corresponding Maximum A Posteriori (MAP) solution is equivalent to minimizing our energy function. The MAP estimation can be efficiently performed by minimization via graph cuts [14], where the data term is encoded using our robust local evidence. We show that the smoothness term can be encoded into a discontinuity-preserving metric, thus allowing the more efficient α-expansion [5] for rapid convergence to the global minimum in a strong sense, instead of the slower swap move [5] in a pairwise MRF. In contrast to [13], while the energy function we minimize is regular, we treat the noisy photometric data *asymmetrically* to infer a dense and unbiased set. On the other hand, like [13], smoothness is enforced while discontinuities are preserved.

The organization of this paper is as follows. Section 2 reviews related work. Section 3 reviews the MRF model for photometric stereo, leading to the discussion of the data term and smoothness term in the graph cut formulation. In section 4, we describe the inference of dense and unbiased set from the noisy observations to model the data term. In section 5, we define our energy function to be minimized by graph cuts, where the smoothness term is detailed. Our graph cut approach is justified by comparison with convex minimization in section 6. Extensive experimental results are presented in section 7. Finally, conclusions are drawn in section 8.

2 Related work

Woodham [23] first introduced photometric stereo for Lambertian surfaces assuming known albedos. Three images are used to solve the reflectance equation for recovering surface gradients p, q and albedo ρ of a Lambertian surface: $R(p,q) = \rho \frac{l_x p + l_y q + l_z}{\sqrt{1+p^2+q^2}}$, where $p = \frac{\partial z}{\partial x}$, $q = \frac{\partial z}{\partial y}$ are surface gradients, $\vec{l} = [l_x \ l_y \ l_z]$ is the unit light direction. This method is pixel-based and hence noise sensitive. Over the past decade, many approaches have been proposed to solve the problem more robustly.

Four images. Coleman and Jain [6] used four images to compute four albedo values at each pixel, using four combinations of three light sources. Presence of specular highlight will make the computed albedos not identical, indicating that some measurement should be excluded. In [18], four images were used. Barskey and Petrou [3] showed that [6] is still problematic if shadows are present, and generalized [6] to handle color images. Neighborhood information may not be adequately considered by these methods.

Reference objects. In [9], a reference object was used to perform photometric stereo. Isotropic materials were assumed. In this approach, outgoing radiance function for all directions are tabulated to obtain an empirical reflectance model. Hertzmann and Seitz [7] used a similar technique, and presented an approach to compute surface orientations and reflectance properties. They made use of orientation consistency to establish correspondence between the unknown object and a known reference object. In many cases, however, obtaining a reference object for correspondence can be very difficult, and in [7] a simplified model was used when such object is unavailable.

Analytic models. By considering diffuse and non-Lambertian surfaces, Tagare and deFigueriredo [19] developed a theory on m-lobed reflective map to solve the problem. Kay and Caelly [12] extended [19] and applied nonlinear regression to a larger number of input images. Solomon and Ikeuchi [18] extended [6] by separating the object into different areas. The Torrance-Sparrow model was then used to compute the surface roughness. Nayar, Ikeuchi and Kanade [17] used a hybrid reflectance model (Torrance-Sparrow and Beckmann-Spizzichino), and recovered not only the surface gradients but also parameters of the reflectance model. In these approaches, the models used are usually somewhat complex, and more parameters need to be estimated.

To our knowledge, there is no previous work on using energy minimization via graph cut to solve the dense photometric stereo problem, despite the desirable properties of a fast and simple implementation with a theoretical guarantee (in a strong sense) in [14]. The use of a dense set of photometric stereo data (> 1000) has not been extensively explored, possibly due to the difficulty in acquiring hundreds of accurate photometric images. An earlier work [16] investigated two algorithms: parallel and cascade photometric stereo for surface reconstruction which use a larger number of images. A related work using one image, that is, shape from shading, was reported in [11], where the problem was solved via graph cuts, by combining local estimation based on local intensities and global energy minimization.

3 The MRF model for photometric stereo

Let $f = \{\mathbf{N}_1, \mathbf{N}_2, \cdots, \mathbf{N}_D\}$ be the pixel-wise normal configuration of the scene, given a set of photometric images $I = \{I_1, I_2, \cdots, I_K\}$ each of dimension D. Following [20] where a MRF model on disparity for stereo is presented, the MRF model for photometric stereo on pixel-wise normal is similar:

$$P(f|I) \propto \prod_p \varphi(\mathbf{N}_p, I_p) \prod_{(p,q)} \phi(\mathbf{N}_p, \mathbf{N}_q) \qquad (1)$$

where φ is a matching cost function associated with corresponding pixels in the captured data. (p,q) is a pair of neighboring pixels (usually first order neighborhood if pairwise MRF is considered). ϕ is a compatibility function between neighboring normals. The best probability is given

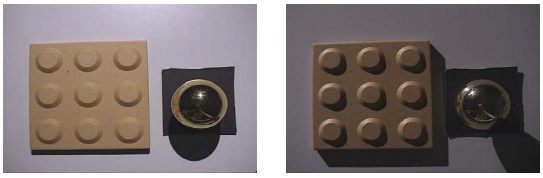


Figure 2: Two views of the experimental set-up under different illumination.

by the MAP solution of the above equation. If we take logarithm of Eqn. (1) it becomes

$$E(f) = \sum_p -\log \varphi(\mathbf{N}_p, I_p) + \sum_{(p,q)} -\log \phi(\mathbf{N}_p, \mathbf{N}_q) \quad (2)$$

$$E(f) = \sum_p D(\mathbf{N}_p, I_p) + \sum_{(p,q)} V(\mathbf{N}_p, \mathbf{N}_q) \quad (3)$$

where the functions D and V are energy functions to be minimized by graph cuts under certain necessary conditions. D and V are called the data term and the smoothness term in graph cuts, which relate to the matching cost and compatibility function of the corresponding MRF model respectively.

4 A simple system

We first describe a very simple system. Typical captured images are very noisy (Figure 1). By resampling dense and noisy observation to infer a uniform set, we can perform normal plane fitting (section 4.4) to obtain a robust data term estimate.

Our system is inspired by [7] where a reference object of known geometry is used to find out surface normals of the target object. They performed BRDF response matching based on the orientation-consistency cue, where the specular highlight implicitly gives surface normal direction. The reference object should be similar to the target object in material. In contrast to [7], we explicitly use the specular highlight to estimate light directions, and then use the estimated light direction to obtain rough initial surface normals based on a simple reflectance model. These normals will be encoded into the data term in our energy minimization. No reference object of similar material is used.

4.1 Estimation of light direction

We need acceptable but not exact light directions. Our method is simple. Shown in Figure 2 is our experimental set-up, where two views of the object and a shiny sphere under different illumination are depicted.

We place a highly reflective sphere next to the object. A DV camera is used to capture a sequence of images by changing the direction of light source. The auto-exposure function of the DV was turned off during data capturing. We tried to hold the spotlight at a constant distance as much as possible to maintain a constant irradiance on the object, but some small changes are unavoidable and so our images suffer various degrees of light intensity attenuation. To capture sufficient directions, it is evitable that shadows of the wire, the camera tripod and the camera were cast to the object.

The mirror reflection on the sphere helps us to estimate lighting direction, which should be a very bright but small point. By searching for the maximum intensity, we can readily localize the point of mirror reflection. Since we know the geometry of the sphere and the viewing direction (assumed to be orthographic), by the law of pure mirror reflection, we can estimate the lighting direction by:

$$\mathbf{L} = 2\mathbf{N}(\mathbf{N} \cdot \mathbf{V}) - \mathbf{V} \quad (4)$$

where $\mathbf{N}$ is the surface normal at the brightest pixel (a, b), $\mathbf{V} = (0, 0, 1)^T$ and $\mathbf{L}$ is the lighting direction. $\mathbf{N}$ can be determined given (a, b), the projection of the sphere center (c_x, c_y), and the projection of radius r. Under orthographic projection, we measure c_x, c_y and r directly on any captured image.

To minimize the error caused by reflections not due to the light source (e.g. from the table where the object and the sphere are placed), we have to limit the search space by considering only the pixels (x, y) satisfying $(x - c_x)^2 + (y - c_y)^2 < r^2 - r^2 \cos(\frac{\pi}{4}) - \epsilon$ where $\epsilon > 0$ is a small constant to offset the small error caused by r, c_x, c_y. Using this condition, all light coming from the direction in the lower hemisphere will be automatically ignored.

4.2 Uniform resampling

The noisy data manually acquired by the above setup corresponds to an unevenly scattered set on the light direction sphere, which will produce undesirable biases to the result. To produce uniform sampling of light directions, we start with an icosahedron, and perform subdivision on each face 4 times recursively [2]. It produces a total of 1313 points uniformly sampled on the light direction sphere. Let us suppose that the object is located at the center of the light direction sphere. Ideally, we want to illuminate the object along the line joining the center and one of the vertices of the subdivided icosahedron, such that uniform distribution can be achieved. In practice, we seek the nearest light direction $\mathbf{L}_o$ at one vertex in the subdivided icosahedron for each captured light direction $\mathbf{L}_i$, and interpolate the image I_o at $\mathbf{L}_o$ by $I_o(x, y) = \sum_{i \in \mathcal{V}} \frac{\mathbf{L}_o \cdot \mathbf{L}_i}{\sum_{i \in \mathcal{V}} \mathbf{L}_o \cdot \mathbf{L}_i} I_i(x, y)$, where $\mathcal{V}$ is a set of indices to the captured light direction closest to $\mathbf{L}_o$. Uniform resampling is thus achieved.

4.3 Denominator image

Complex objects contain textures and spatially varying albedos. By adopting a simple model, we choose an image to cancel out the surface albedo by producing *ratio images*, which will be described in the next section. Here we first

describe the automatic selection of the denominator image, which should be least affected by shadows and highlight. The simple model we use is: each pixel intensity satisfies $\rho(\mathbf{N} \cdot \mathbf{L})$. We show in the result section that this model can be deviated significantly in our robust system.

To choose a suitable image, we make use of the following simple ranking method that works for all our examples: we stack the resampled images into a space-time volume. For each pixel location (x, y), we sort the corresponding pixel intensities over time. Hence, for each pixel in the resampled image, we know its intensity rank over time. Since intensities adversely affected by shadows and specular lightlight will go to the two extremes of the sorted list, for each location (x, y), if the intensity rank at (x, y) is greater than the median and smaller than some upper bound, it has a high probability to be both shadow-free and highlight-free.

Based on the above observation, for each resampled image i, we count the number of pixels whose intensity rank satisfies $rank > L$, where $L \geq$ median. Let k_L^i be the total number of pixels in image i satisfying this condition, $\overline{r}_L^i$ be the mean rank among the pixels in image i that satisfies this condition. The denominator image is defined to be one with 1) maximum k_L and 2) $\overline{r}_L$ lower than some threshold H. Currently, we set L and H to be the 70th and 90th percentiles respectively in all experiments.

4.4 Local normal estimation by ratio images

After uniform resampling and obtaining a denominator image, we are ready to produce an initial normal for each pixel, which is needed for defining our energy functions. After this local estimation, we will use global optimization via graph cuts to aggregate neighborhood support.

Suppose that the object is nearly Lambertian. Then, it can be described by $\rho(\mathbf{N} \cdot \mathbf{L})$, where ρ is the surface albedo, $\mathbf{N}$ is the normal at the pixel and $\mathbf{L}$ is the light direction. Note that $\mathbf{N}$ and ρ are the same for each corresponding pixel over time. Let k be the total number of resampled images. To eliminate ρ, we divide $k-1$ resampled images by the denominator image to obtain $k-1$ *ratio images*. Every pixel in an ratio image is expressed by

$$\frac{I_1}{I_2} = \frac{\mathbf{N} \cdot \mathbf{L_1}}{\mathbf{N} \cdot \mathbf{L_2}}$$

By using this description and no less than three ratio images, we are able to estimate the normal at each pixel locally: Define $\mathbf{N} = (x, y, z)^T$,$\mathbf{L_1} = (l_{1x}, l_{1y}, l_{1z})^T$ and $\mathbf{L_2} = (l_{2x}, l_{2y}, l_{2z})^T$. For each of the pixel in a ratio image, we obtain a plane equation

$$Ax + By + Cz = 0 \tag{5}$$

where $A = I_1 l_{2x} - I_2 l_{1x}, B = I_1 l_{2y} - I_2 l_{1y}, C = I_1 l_{2z} - I_2 l_{1z}$ are constants. Given $k - 1 \geq 3$ ratio images, we have $k-1$ such equations for each pixel. We can solve for $(x, y, z)^T$ by singular value decomposition (SVD) which explicitly enforces that $||\mathbf{N}|| = 1$.

5 Minimization via graph cuts

The notations for graph cut minimization used in this section are more complex than that of the introductory section 3. Let $\mathcal{P}$ be the set of all pixels in the $k-1$ resampled image excluding the denominator image, and $\mathcal{D}$ be the set of all pixels in the denominator image. For each pixel location (x, y), our goal is to obtain an optimal $\mathbf{N}(x, y)$. In graph cut, we seek an optimal labeling $f : \mathcal{P} \rightarrow \mathcal{L}$ where $\mathcal{L} = \{\ell_1, \ell_2, \cdots\}$ is a discrete set of labels corresponding to different normal orientations. In our implementation, again, the labels correspond to the vertices on a subdivided icosahedron which guarantees uniform distribution on a sphere [2]. To increase precision, we follow [2] to subdivide each face of an icosahedron recursively in a total of 5 times, so that $|\mathcal{L}| = 5057$. From our experimental results, this gives seamlessly smooth surface normals on a sphere and acceptable running time.

5.1 Energy function

Our energy function for graph cut minimization consists of two terms: $E(f) = E_{data}(f) + E_{smoothness}(f)$.

Data term. Since our input consists of images and lighting directions only, our data term should measure the per-pixel difference between the measured and the estimated ratio images.

Let $\mathbf{N}(w) = (x(w), y(w), z(w))^T$ be the normal indexed by the label $w \in \mathcal{L}$, $L_p = (l_{px}, l_{py}, l_{pz})^T$ be the lighting vector at pixel $p \in \mathcal{P}$, $L_d = (l_{dx}, l_{dy}, l_{dz})^T$ be the lighting vector at pixel $d \in \mathcal{D}$, where p and d are corresponding pixels. Let I_p be the measured intensity at p, and I_d be the measured intensity at d, and $f(p) = f(d)$ be the normal label at p or d.

Ideally, if the surface is Lambertian, $\frac{I_p}{I_d} - \frac{N(f(p)) \cdot L_p}{N(f(p)) \cdot L_d} = 0$ for all pixels p. By rearranging this equation, we have an equation of same form as (5), $A_p x(f(p)) + B_p y(f(p)) + C_p z(f(p)) = 0$, where $A_p = I_p l_{dx} - I_d l_{px}$, $B_p = I_p l_{dy} - I_d l_{py}$ and $C_p = I_p l_{dz} - I_d l_{pz}$. Our goal is to minimize this equation for all pixels p. Thus, our robust data term is defined as follows, by considering all measurements in $\mathcal{P}$:

$$E_{data}(f) = \sum_{p \in \mathcal{P}} D_p(f(p)) \tag{6}$$

$$= \sum_{p \in \mathcal{P}} [A_p x(f(p)) + B_p y(f(p)) + C_p z(f(p))]^2 \tag{7}$$

Smoothness term. On the other hand, the smoothness term should measure the smoothness of object surface while preserving discontinuity. By using the neighborhood information of a pixel, the effect of noise can be reduced. We define a set of neighboring pixels $\mathcal{N}$ to be a set of pixel pairs p_1, $p_2 \in \mathcal{P}$, where $\{p_1, p_2\}$ is the first order neighbors located

in the same resampled image. The corresponding pixel pair of $\{p_1, p_2\}$ in $\mathcal{D}$ is $\{d_1, d_2\}$.

To define the discontinuity-preserving smoothness term, we employ the surface continuity information. The normals estimated in section 4.4 can be regarded as the initial measurement. Let $\tilde{\mathbf{N}}_{p_1}$ and $\tilde{\mathbf{N}}_{p_2}$ be the normals estimated by section 4.4 at pixel p_1 and p_2 respectively. We define smoothness between p_1 and p_2 to be the difference between $\tilde{\mathbf{N}}_{p_1}$ and $\tilde{\mathbf{N}}_{p_2}$, that is:

$$\tilde{S}_{p_1,p_2} = |\tilde{\mathbf{N}}_{p_1} - \tilde{\mathbf{N}}_{p_2}| \tag{8}$$

So, when the surface is smooth, $\tilde{S}_{p_1,p_2}$ is small and vice versa in the presence of surface orientation discontinuities. Similarly, we measure the smoothness in global minimization by

$$S_{p_1,p_2}(f) = |\mathbf{N}(f(p_1)) - \mathbf{N}(f(p_2))| \tag{9}$$

When $\tilde{S}_{p_1,p_2}$ is small, S_{p_1,p_2} has high probability to be small. Thus, we define our smoothness term as follow:

$$\begin{aligned} & E_{smoothness}(f) \\ = & \sum_{\{p_1,p_2\}\in\mathcal{N}} V_{p_1,p_2}(f(p_1), f(p_2)) && (10) \\ = & \lambda \sum_{\{p_1,p_2\}\in\mathcal{N}} (1+\epsilon-\exp(-\frac{2-\tilde{S}_{p_1,p_2}}{\sigma^2}))S_{p_1,p_2}(f) && (11) \\ = & \lambda \sum_{\{p_1,p_2\}\in\mathcal{N}} K_{p_1,p_2} S_{p_1,p_2}(f) && (12) \end{aligned}$$

where $K_{p_1,p_2} = (1+\epsilon-\exp(-\frac{2-\tilde{S}_{p_1,p_2}}{\sigma^2}))$, λ is the weighting of the smoothness term, σ controls the smoothness uncertainty and $\epsilon > 0$ is a small constant.

5.2 Graph construction

To perform multi-labeling minimization, the expansion move algorithm [14] is one suitable choice. Here, we have a quick review on this algorithm:

α-expansion. For each iteration, we simply select a normal direction label $\alpha \in \mathcal{L}$, and then find the best configuration within this α-expansion move. If this configuration reduces the user-defined energy, the process is repeated. Else, if there is no α that decreases the energy, we are done.

According to [14], the user-defined energy function has to be regular and thus graph-representable so that it can be minimized via graph cut (in a strong sense). This is also true for $|\mathcal{L}|$-label configuration if α-expansion is employed. More precisely, for our $|\mathcal{L}|$-label case, the energy function has to be regular for each α displacement. In this section, we will prove that our energy function E is regular.

For any class $\mathcal{F}^2$ function of the form defined in [14]:

$$E(x_1, ..., x_n) = \sum_i E^i(x_i) + \sum_{i<j} E^{i,j}(x_i, x_j) \tag{13}$$

where $\{x_i | i = 1, ..., n\}$ and $x_i \in \{0, 1\}$ is a set of binary-valued variables. E is regular if and only if

$$E^{i,j}(0,0) + E^{i,j}(1,1) \leq E^{i,j}(0,1) + E^{i,j}(1,0) \tag{14}$$

From [14], it is known that any functions of one variable are regular and hence the data term E_{data} is regular. Therefore, it remains to show that the smoothness term $E_{smoothness}$ satisfies (14) within a move. We prove the following claim on V, which makes E regular. This claim also allows for the more efficient α-expansion which runs in $\Theta(|\mathcal{L}|)$ time [14].

Claim: V_{p_1,p_2} is a metric.

The proof is as follows. In order that V is a metric, for any label $a_1, a_2, a_3 \in \mathcal{L}$, the following three conditions have to be satisfied:

$$\begin{aligned} V(a_1, a_2) = 0 & \Leftrightarrow a_1 = a_2 \\ V(a_1, a_2) & = V(a_2, a_1) \geq 0 \\ V(a_1, a_2) & \leq V(a_1, a_3) + V(a_3, a_2) \end{aligned}$$

Since the first two conditions are trivially true for our $E_{smoothness}$, we shall focus on the third condition here. For any pair of p_1 and p_2, we write:

$$V_{p_1,p_2}(a_1, a_3) + V_{p_1,p_2}(a_3, a_2) - V_{p_1,p_2}(a_1, a_2) \tag{15}$$

$$\begin{aligned} = & K_{p_1,p_2}\{|\mathbf{N}(a_1) - \mathbf{N}(a_3)| + \\ & |\mathbf{N}(a_3) - \mathbf{N}(a_2)| - |\mathbf{N}(a_1) - \mathbf{N}(a_2)|\} \geq 0 \end{aligned} \tag{16}$$

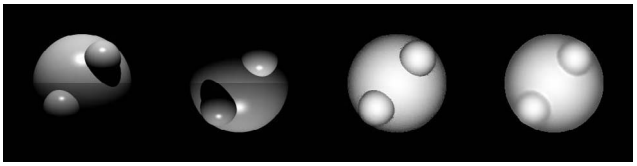
Note that $\mathbf{N}(a_1) - \mathbf{N}(a_3), \mathbf{N}(a_3) - \mathbf{N}(a_2)$ and $\mathbf{N}(a_1) - \mathbf{N}(a_2)$ are three vectors projected onto the same plane defined by the points $\mathbf{N}(a_1)$, $\mathbf{N}(a_2)$ and $\mathbf{N}(a_3)$, which form a triangle on the plane. By the triangle inequality, (16) must not be less than zero, and hence the third metric condition holds. □

Since V_{p_1,p_2} is a metric, $V_{p_1,p_2}(\alpha, \alpha) = 0$ and $V_{p_1,p_2}(f(p_1), f(p_2)) \leq V_{p_1,p_2}(f(p_1), \alpha) + V_{p_1,p_2} f(\alpha, f(p_2))$, the smoothness term $E_{smoothness}$ is regular [14]. To minimize our energy function in each α displacement, we can construct a graph by using [14], and then apply the max-flow algorithm.

6 Why minimization via graph cut?

In this section, we justify the use of graph cut instead of using conventional convex minimization. In fact, if we replace the term $S_{p_1,p_2}(f)$ in (9) in the smoothness term (11) by $[\mathbf{N}(f(p_1)) - \mathbf{N}(f(p_2))]^2$, our energy function becomes convex and therefore can be minimized by standard techniques. It *seems* that we can directly estimate the continuous surface normals, rather than using the combinatorial optimization by graph cuts on a set of discrete normal labels. Our convex energy function has the following form:

$$E(f) = E_{data}(f) + E_{smoothness}(f) \tag{17}$$



(a) (b) (c) (d)

Figure 3: (a)–(b) Two views of the input images. (c) Normals reconstructed by graph cut. (d) Normals reconstructed by convex minimization. Note that discontinuity is not preserved in (d). For ease of visualizing the recovered normals, in (c) and (d), we treat the object as Lambertian and display $\mathbf{N} \cdot \mathbf{L}$, where $\mathbf{L} = (0\ 0\ 1)^T$ and $\mathbf{N}$ is the recovered normal at a pixel.

$$= \sum_{p \in \mathcal{P}} [A_p x(p) + B_p y(p) + C_p z(p)]^2$$

$$+ \lambda \sum_{\{p_1, p_2\} \in \mathcal{N}} K_{p_1,p_2} [\mathbf{N}(p_1) - \mathbf{N}(p_2)]^2 \quad (18)$$

where $f : \mathcal{P} \rightarrow R^3$, $f(p) = \mathbf{N}(p) = (x(p), y(p), z(p))^T$ is the normal at p which is a 3-vector variable. It is known that energy function in this form can be solved by some iterative minimization algorithm such as gradient decent and Levenberg-Marquardt method to obtain a global minimum. However, the unity constraint, that is, the fact that surface normals have to be unit vectors, has to be enforced.

One way is to enforce the unity constraint is to rewrite every $\mathbf{N}(p)$ in the energy function to $\frac{\mathbf{N}(p)}{||\mathbf{N}(p)||}$. However, it complicates the energy function. The other alternative is to add the following term to the energy function:

$$E_{unit}(f) = \gamma \sum_{p \in \mathcal{P}} (\mathbf{N}(p) \cdot \mathbf{N}(p) - 1)^2 \quad (19)$$

where γ is the weighting of E_{unit} in the energy function. Thus E_{unit} introduces a soft constraint to the function, and the resulting normals may not be exactly unit vectors.

We may enforce the unit vector constraint by an iterative method that optimizes and normalizes each of the normal one-by-one by fixing other normals. Using this method, by adding up (19) and (18) and taking the first derivative to the resulting energy function, we obtain a system of equations of the following form: $\mathbf{A}\mathbf{N}(p) = \lambda \vec{b}$ where A is a 3×3 symmetric and invertible matrix, and $\vec{b}$ is a non-zero 3-vector. Note that the λ is multiplied to the smoothness term. By solving this linear system, we can obtain $\mathbf{N}(p)$. However, if we normalize $\mathbf{N}(p)$, it means that we are scaling λ *locally* at each pixel, while λ relates to the smoothness term in the *global* energy function. Therefore, the result becomes highly unpredictable. Figure 3 compares the result on a synthetic object obtained by graph cut and this method. Problem exists around the location of orientation discontinuities.

Therefore, in conclusion, to minimize E, graph cut is the good choice because the unity constraint comes for free by $\mathcal{L}$. Graph cut also preserves discontinuity by using a simple but effective energy function. We expect that belief propagation using MAP or MMSE estimators should produce comparable result, because it describes the same MRF [20].

7 Results

We implemented our MRF photometric stereo model that employs the inherent stability and robustness provided by our densely resampled and unbiased but noisy data, and run extensive experiments to evaluate our approach. In all cases, accurate normals can be recovered. We also show some surfaces we reconstructed from the recovered normals at novel viewpoints using [15]. We tested very difficult objects and scenes with a lot of highlight and shadows, and even objects with transparency and other violations to our simple assumed model to demonstrate the robustness of our method. For visualization, the normal $\mathbf{N}$ recovered at each pixel is displayed by $(\mathbf{N} \cdot \mathbf{L})$ where $\mathbf{L}$ is a synthetic light.

Complex patterns with discontinuities. Note the high level of details we can achieve in our reconstruction in Figure 1, where the cloud, snail and mushroom and other complex patterns are faithfully preserved in the normal reconstruction, despite the presence of cast shadows and hightlights in various areas in the entire data set. The smooth surface and underlying surface orientation discontinities are faithfully restored. We also show the result of 3D texture mapping on a cylinder by using our reconstructed surface and normals for this object (Figure 1 (f)). Figure 4 shows other results in this category .

Complex geometry with spatially-varying albedos. Figure 5 shows a difficult example with very complex albedos and geometry (e.g. the steering wheel, the string, and many surface orientation discontinuities). Although severe shadows and highlight exist, we can produce good albedos and surface normal map, except for the headlights which are black in color. The albedo image is obtained by dividing one of the input image by the $\mathbf{N} \cdot \mathbf{L}$ image with the corresponding $\mathbf{L}$.

Complex objects. We show an array of results in Figure 7. The geometry and albedos of the teapot are complex. Our method can faithfully reconstruct the normal directions and shape of the teapot, even the small air hole on the lid, while rejecting all noises caused by the complex patterns. Similar results are obtained for the teddy bear, where problems occur only at the black eyebrows and nose. For the nose, the errors only show up along its boundary. The rope has spatially-varying surface mesostructures, but is faithfully reconstructed.

Complex objects with transparency. Finally, Figure 6 tested our system to the limit. The toy we bought is contained inside an open paper box, which cast a lot of shadows when the spotlight is illuminated on either sides. The toy is wrapped inside a transparent plastic container. So when it is illuminated at other directions, a lot of highlight is produced. Surface orientation discontinuities are ubiquitous in

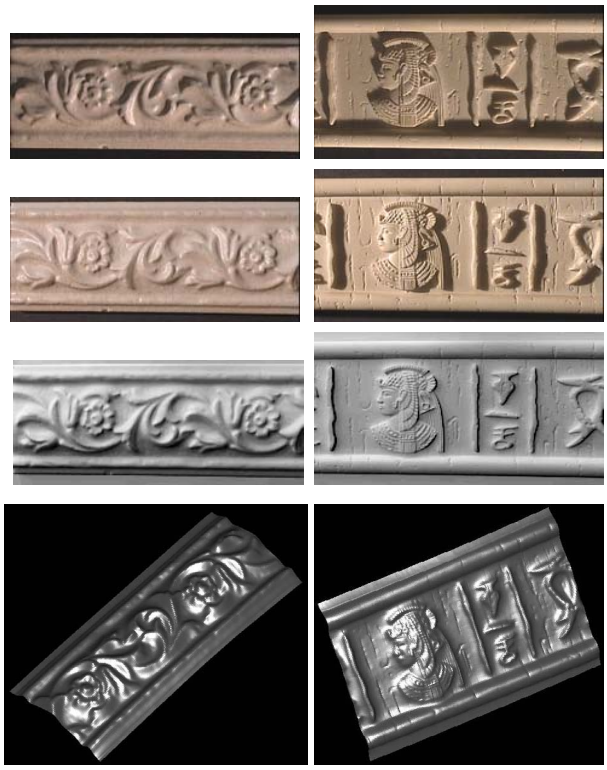


Figure 4: Results on two tiles. The top two rows show two typical images we captured. The reconstructed normals are shown here as $\mathbf{N} \cdot \mathbf{L}$ with $\mathbf{L} = (-\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})^T$. The last row shows the reconstructed surfaces at novel viewpoints.

the object. It is very tedious to choose the right frames from the 3000+ frames we captured to perform sparse photometric stereo and unbiased statistics is not guaranteed. Our simple system which utilizes dense, uniform but noisy information can effectively deal with these problems. The surface normals we recovered are very reasonable under this complex situation. Some errors are at the labels, which are due to the severe violation of our simple assumption that ratio images can divide out albedos here.

8 Conclusion

Shadows, highlight, complex surface reflectance, spatially-varying albedos, complex geometry and inaccurate light directions are major issues confronted by all photometric stereo reconstruction systems. By employing a simple reflectance model, we translate the dense, uniform but inaccurate observation into a robust estimate for the matching cost, which is encoded into a robust data term. It operates with the smoothness term in energy minimization for further normal refinement while preserving discontinuities. We compare our approach with conventional convex minimization. No reference object of similar material is needed. A mirror sphere is used to capture light directions. We have presented very good results in spite of many model violations. In the future, we will investigate the use of a better model to account for complex albedos in reconstruction.

Figure 5: The top row shows two typical views of our noisy data, where shadows and highlight are present on the complex object. The bottom left shows the recovered albedo. The bottom right shows the recovered surface normals displayed using $\mathbf{N} \cdot \mathbf{L}$ with $\mathbf{L} = (\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})^T$. Please see supplementary video [1].

References

[1] http://www.cs.ust.hk/∼pang/cvpr05/photo.html.

[2] D.H. Ballard and C.M. Brown. Computer vision. In *Prentice Hall*, 1982.

[3] S. Barsky and M. Petrou. The 4-source photometric stereo technique for three-dimensional surfaces in the presence of highlights and shadows. *PAMI*, 25(10):1239–1252, October 2003.

[4] R. Basri and D.W. Jacobs. Photometric stereo with general, unknown lighting. In *CVPR01*, pages II:374–381, 2001.

[5] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *PAMI*, 23(11):1222–1239, November 2001.

[6] E.N. Coleman, Jr. and R. Jain. Obtaining 3-dimensional shape of textured and specular surfaces using four-source photometry. *CGIP*, 18(4):309–328, April 1982.

[7] A. Hertzmann and S.M. Seitz. Shape and materials by example: a photometric stereo approach. In *CVPR03*, pages I: 533–540, 2003.

[8] B.K.P. Horn. *Robot Vision*. McGraw-Hill, 1986.

[9] B.K.P. Horn, R.J. Woodham, and W.M. Silver. Determining shape and reflectance using multiple images. In *MIT AI Memo*, 1978.

[10] http://cvc.yale.edu/projects/yalefaces/yalefaces.html.

[11] Y. Ju, K. Lee, and S.U. Lee. Shape from shading using graph cuts. In *ICIP03*, pages I: 421–424, 2003.

[12] G. Kay and T. Caelly. Estimating the parameters of an illumination model using photometric stereo. *GMIP*, 57(5):365–388, 1995.

[13] V. Kolmogorov and R. Zabih. Multi-camera scene reconstruction via graph cuts. In *ECCV02*, page III: 82 ff., 2002.

[14] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts? *PAMI*, 26(2):147–159, February 2004.

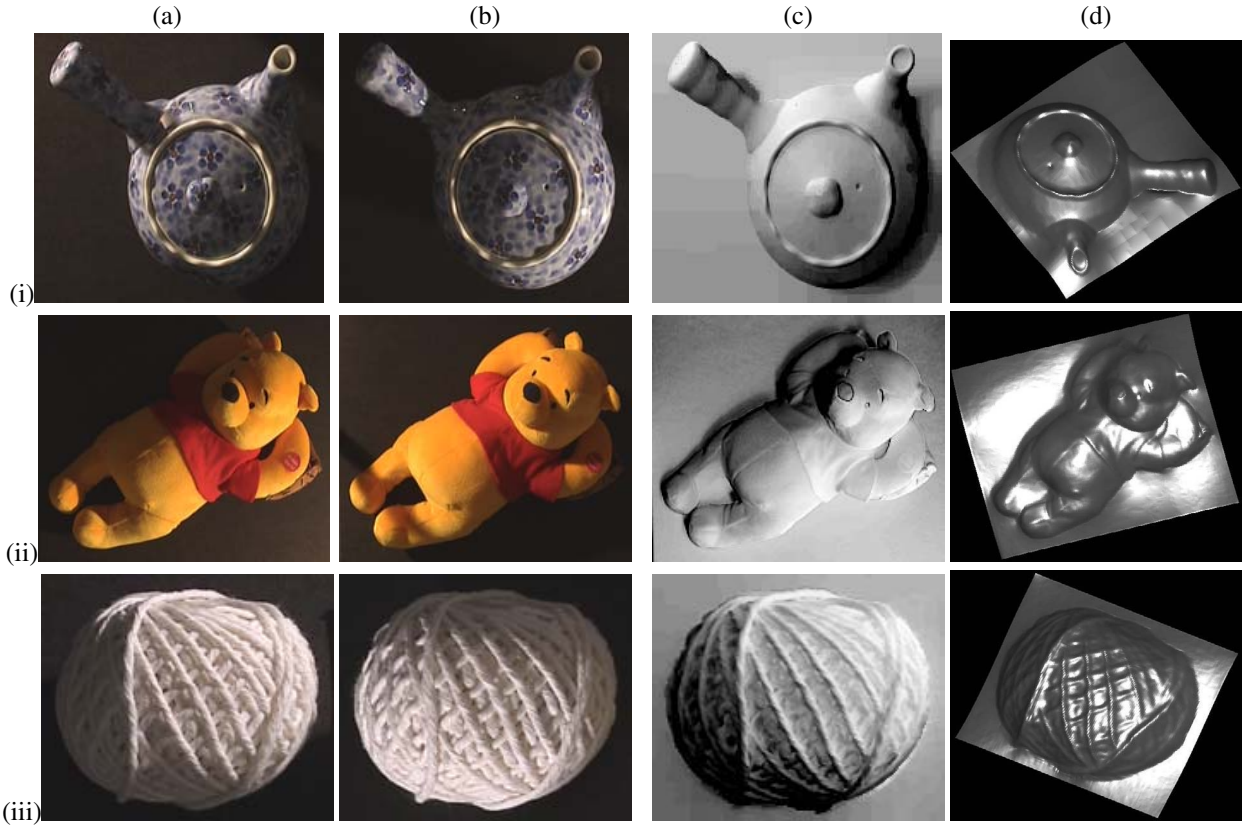


Figure 7: (a)-(b) show two typical captured images of three objects with complex geometry, texture, and/or mesostructures and severe shadows. (c) shows our recovered normals $\mathbf{N}$, each of them is displayed as $\mathbf{N}\cdot\mathbf{L}$ with $\mathbf{L} = (-\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})^T$ for (i), $\mathbf{L} = (\frac{1}{\sqrt{3}}, -\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})^T$ for (ii) and $\mathbf{L} = (\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})^T$ for (iii) respectively. (d) shows the reconstructed surfaces rendered at novel viewpoints.

[15] Peter Kovesi. Shapelets correlated with surface normals produce surfaces. Technical report.

[16] K.M. Lee and C.C.J. Kuo. Shape reconstruction from photometric stereo. In *CVPR92*, pages 479–484, 1992.

[17] S.K. Nayar, K. Ikeuchi, and T. Kanade. Determining shape and reflectance of hybrid surfaces by photometric sampling. *IEEE Trans. on Robotics and Automation*, 6(4):418–431, 1990.

[18] F. Solomon and K. Ikeuchi. Extracting the shape and roughness of specular lobe objects using four light photometric stereo. *PAMI*, 18(4):449–454, April 1996.

[19] H.D. Tagare and R.J.P. deFigueiredo. A theory of photometric stereo for a class of diffuse non-lambertian surfaces. *PAMI*, 13(2):133–152, February 1991.

[20] M.F. Tappen and W.T. Freeman. Comparison of graph cuts with belief propagation for stereo, using identical mrf parameters. In *ICCV03*, pages 900–907, 2003.

[21] Xin Tong, Jingdan Zhang, Ligang Liu, Xi Wang, Baining Guo, and Heung-Yeung Shum. Synthesis of bidirectional texture functions on arbitrary surfaces. In *ACM SIGGRAPH*, pages 665–672, 2002.

[22] M. Alex O. Vasilescu and Demetri Terzopoulos. Tensortextures: multilinear image-based rendering. *ACM Trans. Graph.*, 23(3):336–342, 2004.

[23] R.J. Woodham. Photometric method for determining surface orientation from multiple images. *OptEng*, 19(1):139–144, January 1980.

[24] R.J. Woodham. Gradient and curvature from the photometric-stereo method, including local confidence estimation. *JOSA-A*, 11(11):3050–3068, November 1994.

Figure 6: Three typical images for this object are shown in the top, where many assumptions used in previous photometric stereo systems are violated: shadows, highlight, transparency, and inter-reflections due to the complex geometry and spatially-varying albedos. The recovered normals $\mathbf{N}$ at each pixel, displayed here as $\mathbf{N}\cdot\mathbf{L}$ with $\mathbf{L} = (\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})^T$, are very reasonable.