Subpixel Photometric Stereo

Ping Tan, Stephen Lin, and Long Quan Senior Member, IEEE
Abstract

Conventional photometric stereo recovers one normal direction per pixel of the input image. This fundamentally limits the scale of recovered geometry to the resolution of the input image, and does not adequately model surfaces with subpixel geometric structures. In this paper, we propose a method to recover subpixel surface geometry by studying the relationship between the subpixel geometry and the reflectance properties of a surface. We first describe a generalized physically-based reflectance model that relates the distribution of surface normals inside each pixel area to its reflectance function. The distribution of surface normals can be computed from the reflectance functions recorded in photometric stereo images. A convexity measure of subpixel geometry structure is also recovered at each pixel, through an analysis of brightness attenuation due to shadowing. Then, we use the recovered distribution of surface normals and the surface convexity to infer subpixel geometric structures on a surface of homogeneous material by spatially arranging the normals among pixels at a higher resolution than that of the input image. We optimize the arrangement of normals using a combination of belief propagation and MCMC based on a minimum description length criterion on 3D textons over the surface. Experiments demonstrate the validity of our approach and show superior geometric resolution for the recovered surfaces.

Index Terms

Photometric stereo, superresolution, 3D textons.

I. INTRODUCTION

From a given viewing direction, the appearance of surface points vary according to their orientation, reflectance, and illumination conditions. With an assumed reflectance, photometric stereo methods utilize this relationship to compute surface normals by examining transformations in image intensities that result from changes in lighting directions. Traditionally, reflectance is assumed to be Lambertian [1], [2], and with calibrated illumination directions, three images are sufficient to recover surface normals and albedos [3].

The reflectance of a real surface, however, often does not adhere to the Lambertian model, and in such cases, conventional photometric stereo may yield poor results. To deal with this problem, methods based on non-Lambertian reflectance models have been proposed. Some techniques utilize a composite reflectance model that consists of Lambertian diffuse reflection plus specular reflection [4]–[6], while others employ physically-based models that account for the effects of fine-scale roughness in surface structure [7], [8]. In all of these approaches, the reflectance of each pixel is assumed to be a function of a single principal normal direction.
In many instances, the surface structure within a pixel exhibits greater complexity, and the resulting reflectance cannot be accurately expressed in terms of a single normal. Conventionally in physically based reflectance modeling, an object surface consists of many fine-scale microfacets, and surface reflectance depends upon the distribution of normal directions among these microfacets. This distribution is typically considered to be centered at a single principal direction (e.g., [9], [10]). As shown in the left of Fig. 1, when a surface is imaged at a resolution coarser than its surface structure, multiple disparate principal normal directions may exist within each pixel. Since previous photometric stereo methods compute a normal map at the same resolution as the input images, they are unable to recover this subpixel geometric structure.

In this work, we present a technique for estimating the geometry of a uniform-material surface at a resolution higher than that of the input images. The proposed method recovers for each pixel a general distribution of microfacet normals and its surface convexity from an ample number of photometric stereo images, and then estimates a spatial arrangement of this normal distribution in a higher resolution image. For normal recovery, we perform photometric stereo with a reflectance model formulated using a general representation of normal distributions. An Expectation-Maximization approach is presented to improve robustness in estimating the parameters of these general distributions. The convexity of subpixel geometric structure at each pixel is estimated using a simple convexity metric based on the variation of shadowing attenuation at grazing and non-grazing lighting angles.
With the recovered normal distributions and surface convexity of each pixel, enhanced resolution of surface geometry is computed by dividing the distribution according to the targeted level of enhancement, e.g., four sub-distributions for a 2x2 enhancement of photometric stereo. After partitioning the distribution, their arrangement in the higher resolution normal map is formulated from constraints that favor consistency of geometric structure over a surface. Consistency is evaluated in terms of per-pixel convexity and surface integrability, as well as simplicity of surface description with respect to 3D textons, which is motivated by the minimum description length principle [11] and the observation that a surface is generally composed of only a small number of perceptually distinct local structures [12]. To solve this complicated arrangement problem, we utilize the Belief Propagation algorithm [13], [14] to compute an initial solution from a graphical model that represents integrability constraints. Starting from this initial solution, a Markov Chain Monte Carlo (MCMC) [15] method with formulated proposal functions is used to find an optimal arrangement that accounts for complexity of the surface description.

This approach enhances resolution differently from image superresolution methods [16], [17] and stereo techniques in that subpixel viewpoint displacements are not used to obtain variations in spatial sampling. In photometric stereo, all images are captured at a fixed viewpoint, and the proposed technique estimates higher spatial resolution based on superresolution recovery of surface normals and constraints on surface structure. The described approach also differs from learning-based hallucination methods [18], [19] that utilize a training set of high resolution / low resolution image pairs to infer enhanced resolution. Unlike image hallucination methods, our technique is able to recover partial information from the input for resolution enhancement, in the form of the actual normal distribution within a low resolution pixel. Therefore, our method need not fully conjecture on the high resolution data, which is difficult to do by hallucination because reliable training databases are challenging to construct for general geometric structure. Rather, it infers only the arrangement of known surface normal information. With this approach, fine-scale surface detail that is missed in conventional photometric stereo can be revealed.

II. REFLECTANCE AND SUBPIXEL GEOMETRIC STRUCTURE

We first describe a generalized reflectance model that relates surface reflectance to its subpixel geometric structures. We then discuss how the convexity of subpixel geometry affects pixel brightness due to shadowing attenuation. Recovery of these subpixel level geometric attributes will later
be described in Sec. III. And the recovered attributes are later used to compute surface shape at enhanced resolution.

A. Microfacet Normal Distribution

In physically-based reflectance modeling [10], [20], [21], a surface is typically modeled as a collection of tiny flat faces, called microfacets. The overall reflectance of a surface area imaged within a pixel is therefore an aggregate effect of this microfacet collection, which is generally described by the distribution of their normal directions. For an arbitrary microfacet normal distribution $p(n)$, reflectance may be physically represented by a model proposed in [21]:

$$\rho(l, v) = \frac{p(h)F(l \cdot h)}{4K_s(l)K_m(v)}$$

where $l, v$ are unit lighting and viewing directions, and $h$ is their unit bisector. $F$ denotes Fresnel reflectance, and $K_s(l), K_m(v)$ are factors that account for shadowing and masking among microfacets. Since the viewing direction $v$ is fixed during the acquisition of photometric stereo images, $K_m(v)$ is constant for each pixel throughout the image sequence. Since microfacet-based models generally treat microfacets as mirror reflectors, recovery of $p(h)$ gives us $p(n)$.

Generally in reflectance modeling, the collection of microfacet normals within a pixel is considered to be distributed around a principal surface normal direction. As shown in Fig. 1, at low resolution, sub-pixel geometric structure can lead to a complex distribution of normals within a pixel. When the scale of geometric structure is smaller than the image resolution, multiple principal surface normals may exist within a pixel. As a more general representation of normal distributions, we utilize Gaussian mixture models (GMMs), which have long been used to represent general distributions. To facilitate estimation of reflectance parameters, the shadowing term $K_s$ and Fresnel term $F$ are each modeled as constant, as done in numerous reflectance modeling works (e.g., [5], [21], [22]). We will, however, later relax the constant model of $K_s$ to evaluate the convexity of each pixel.

With this simplification and a normal distribution represented by a Gaussian mixture model $G(n) = \sum_{i=1}^{N} \alpha_i g(n; \mu_i, \sigma_i)$, where $\sum_{i=1}^{N} \alpha_i = 1$, $g(n; \mu_i, \sigma_i)$ is a Gaussian function with $\mu_i, \sigma_i$ as its mean and variance, we can express reflectance as

$$\rho(l, v) = \frac{F}{4K_s K_m} \sum_{i=0}^{N} \alpha_i g(h; \mu_i, \sigma_i) = A \cdot G(h)$$

(1)
where $A = \frac{F}{4K_sK_m}$ is a constant. By incorporating this generalized reflectance model into photometric stereo, our method acquires more comprehensive surface normal information within each pixel.

### B. Shadowing between Microfacets

The shadowing and masking of microfacets is influenced mainly by the microfacet spatial arrangement. As discussed in [21], shadowing and masking terms in many cases have less impact on reflectance than the microfacet normal distribution. The effect of shadowing and masking often becomes significant only at grazing incident lighting angles. Since masking terms are fixed for each pixel in a photometric stereo image sequence, we will focus the discussion on shadowing among microfacets.

Fig. 2 shows two different configurations of a given normal distribution. Although the examples on the left and right have identical microfacet normal distributions, their different spatial configurations result in a different proportion of illuminated area for a grazing incident lighting direction. The shadowing attenuations for these two configurations are therefore different. In general, a convex configuration has less shadowing attenuation than a concave structure at grazing incident light directions. As illustrated by the example in Fig. 2, a pixel with a convex structure can be at least half exposed to the lighting direction, while most of a concave pixel will be occluded when its...
surface slope is steeper than the grazing illumination angle. We will utilize this general difference in microfacet shadowing to infer surface convexity in Sec. III, and this information will later be used to help reduce geometric ambiguity of microfacet normal arrangements in Sec. IV.

III. Recovery of Subpixel Geometric Attributes

From the described relationship between reflectance and subpixel structure, we estimate the microfacet normal distribution and surface convexity at each pixel of the photometric stereo images. The microfacet normal distribution is computed by fitting the generalized reflectance model described in Eq. (1) to the observed pixel intensities with an EM-based algorithm. The pixel-wise convexity is computed by comparing the shadowing attenuation at grazing and non-grazing lighting directions.

A. Recovery of Normal Distributions

We first employ the generalized reflectance model in photometric stereo to recover a general normal distribution for each pixel. Due to the complexity of a general normal distribution, it is non-trivial to determine the parameters of this reflectance model from photometric stereo images. In photometric stereo, a set of \( K \) images containing reflectance data \( \{O_k; 1 \leq k \leq K\} \) is measured under different lighting conditions \( \{l_k; 1 \leq k \leq K\} \) and a fixed viewing direction \( v \). From this data, parameters of the Gaussian mixture normal distribution could in principle be estimated at each pixel by general non-linear least squares fitting:

\[
\Theta = \arg \min \sum_{k=1}^{K} ||\rho(l_k, v) - \hat{O}_k||^2
\]

\[
= \arg \min \sum_{k=1}^{K} ||A \sum_{i=0}^{N} \alpha_i g(h_k; \mu_i, \sigma_i) - \hat{O}_k||^2,
\]

where \( \Theta = \{A, (\alpha_i, \mu_i, \sigma_i); 1 \leq i \leq N\} \) signifies the reflectance parameters of an \( N \)-Gaussian GMM, and \( \hat{O}_k \) is the reflectance value\(^1\). However, as described in [22], due to the high nonlinearity

\(^1\hat{O}_k = O_k/(r_k n \cdot l_k)\), where \( n \) is the single principal normal direction in a pixel, which is computed by conventional photometric stereo before applying our subpixel algorithm. \( r_k \) is the radiant intensity for the \( k \)-th lighting condition. In our work, \( r_k \) is a constant \( r \) for all images, which is achieved by using a fixed distance between lights and object. In practice, this \( r \) can further be set to 1, since only relative values of lighting intensity and surface reflectivity are recoverable, and thus can be rescaled.
of reflectance functions, fitting a model with more than two lobes by general non-linear least squares is rather unstable and gives unreliable results.

To deal with this issue, we first assume $A$ is known and estimate the other parameters. As shown in the following derivation, this estimation is actually independent of $A$. $A$ can be later computed with the determined values of the other parameters. Now, $\rho(l, v)/A = G(h)$ is a probability distribution function (pdf) defined on a hemisphere with respect to bisector direction $h$. From the mirror reflection at microfacets and the pixel intensity measurements of different known light directions, we have the pdf value at certain $h_k$ as $G(h_k) = \hat{O}_k/A$. We need to estimate the pdf function according to these values. If we consider these measurements as samples, a sample at each corresponding bisector direction on the hemisphere, we can compute a distribution from these samples. If each sample is treated equally, the computed distribution will be $t(h)$, which is the distribution of bisector directions on the hemisphere. With importance sampling [15], the computed distribution becomes $G(h)$ when weighting each sample by $G(h)/t(h)$. In this way, we can estimate $G(h)$ from a set of weighted samples. Estimation of a pdf in terms of a Gaussian mixture model from a set of weighted samples is a well-studied problem and can be robustly computed by the Expectation-Maximization (EM) algorithm [23], [24]. With a generalized reflectance model that represents normal distributions with a GMM, we can conveniently utilize this method to recover this detailed surface information.

1) Uniformly Sampled Lighting Directions: Now, let us assume the lighting directions are evenly sampled over the hemisphere. In other words, $t(h)$ is a constant value $t_0$ over the hemisphere. The weight of a sample at $h_k$ should thus be $G(h_k)/t_0 = \hat{O}_k/(At_0)$. For each observation $k$, we instead place a sample of weight $\hat{O}_k = A \cdot G(h_k)$ at direction $h_k$, which we later show will lead to a result equivalent to using weights $G(h_k)/t_0$. To estimate the actual pdf, GMM parameters can be computed according to these weighted samples using the EM algorithm, by iteratively computing the E-step:

$$Ez_{ik} = \alpha_i g(h_k; \mu_i, \sigma_i) / \sum_{j=1}^{N} \alpha_j g(h_k; \mu_j, \sigma_j)$$
Fig. 3. In capturing photometric stereo images, the sampling distribution of lighting directions on the hemisphere will affect the estimation of GMM reflectance parameters.

and the M-step:

\[
\alpha_i = \frac{\sum_{k=1}^{K} \hat{O}_k E z_{ik}}{\sum_{k=1}^{K} \hat{O}_k}
\]

\[
\mu_i = \frac{\sum_{k=1}^{K} \hat{O}_k E z_{ik} h_k}{\sum_{k=1}^{K} \hat{O}_k}
\]

\[
\sigma_i^2 = \frac{\sum_{k=1}^{K} \hat{O}_k E z_{ik} ||h_k - \mu_i||^2}{\sum_{k=1}^{K} \hat{O}_k}
\]

where \(z_{ik}\) are hidden variables, and \(E z_{ik}\) are their expectations. For purposes of resolution enhancement as later described in Sec. IV, we utilize GMMs with Gaussians of equal weight, such that we set \(\alpha_i = \frac{1}{N}\) for an \(N\)-Gaussian GMM.

Since \(A\) and \(t_0\) are constants and \(\hat{O}_k\) appears in both the numerators and denominators of the EM formulas, the use of weights \(\hat{O}_k\) is equivalent to using \(G(h_k)/t_0\) in computing the normal distribution parameters. Intuitively, these GMM parameters, which describe statistical characteristics of a surface, are independent of \(A\), which represent spatial and optical properties, and can be optimized separately. With the computed GMM parameters, \(A\) may be solved by linear least squares optimization:

\[
A = \arg \min \sum_{k=1}^{K} ||\hat{O}_k - AG(h_k)||^2.
\]

Later, we show how this factor can be used to infer per-pixel convexity.
2) Nonuniformly Sampled Lighting Directions: The method discussed above is only valid for uniformly sampled lighting directions on the hemisphere. As illustrated in Fig. 3, if the lighting directions for photometric stereo are sampled nonuniformly, the estimated result will be biased by the distribution of the lighting samples. This is a general problem in model fitting; more densely sampled areas contribute more to the error measure, so the fitting is more focused on those areas. As a result, a solution may achieve lower overall error by slightly improving the fit in densely sampled areas at the expense of relatively large deviations in sparsely sampled areas.

Suppose the pdf of a lighting sample distribution is \( t(h) \). To compensate for non-uniformity in sampling, we weigh the samples at \( h_k \) by \( 1/t(h_k) \) and put a sample of weight \( \hat{O}_k/t(h_k) \) at direction \( h_k \) for each observation. From these weighted samples, the EM algorithm can be used to compute the distribution \( G(\cdot) \). In practice, we compute a Voronoi diagram over the lighting samples on the hemisphere, and set \( t(h_i) \) according to the area of the spherical polygon containing \( h_i \).

The performance of this EM-based parameter estimation is exemplified in Fig. 4, which compares a radiance distribution computed with our generalized model and recovery technique to the observed data from photometric images with 65 sampled illumination directions. The intensity of a single pixel under different lighting directions is visualized as a function defined on a hemisphere, whose value is represented by the radius. Shown in the middle is the result without consideration of bias in the lighting distribution. This distribution is severely biased toward the top of the hemisphere, because of denser sampling at the top during data capture. The distribution on the right is the fitting result where samples are properly weighted with respect to the lighting distribution over the hemisphere. While some deviation can be observed at grazing angles, the fitted radiance distribution gives an approximate 4-Gaussian model of the captured distribution. With greater numbers of Gaussians, closer approximations can be obtained.

B. Recovery of Surface Convexity

According to the discussion in Sec. II, the masking term \( K_m \) of each pixel is constant over the entire sequence of photometric stereo images, while shadowing attenuation \( K_s(l) \) may vary. More importantly, the profile of \( K_s(l) \) differs between pixels with a concave and convex structure. Generally, due to the shadowing between subpixel geometrical structures, concave pixels tend to have less illuminated area than convex pixels for grazing lighting directions, which causes its shadowing attenuation factor to increase at grazing angles. In contrast, convex pixels tend not to
have an appreciable increase in shadowing attenuation. Here, we utilize this difference to infer pixel convexity from its shadowing attenuation profile. To do so, we fix the estimated normal distribution $p(n)$ and compute $K_s(l)$ via the least squares method. However, since the shadowing term $K_s(l)$, masking term $K_m$, and Fresnel term $F$ are multiplied together, we cannot decouple $K_s(l)$ from the other factors easily. Nevertheless, the profile of a convex and concave pixel can be distinguished without computing the exact value of $K_s(l)$. We instead compute the ratio between $K_s(l_{grazing})$ and $K_s(l_{non-grazing})$, where $l_{grazing}$ and $l_{non-grazing}$ are lighting directions with respective polar angles of $80^\circ$ and $60^\circ$ in our implementation. If $K_s(l_{non-grazing})/K_s(l_{grazing})$ is less than a specified threshold, it indicates that shadowing attenuation increases at grazing lighting directions, which implies severe shadowing between microfacets and that the pixel surface is concave. Otherwise, the shadowing attenuation factor does not change significantly with changes in lighting direction, which indicates less shadowing and that the pixel surface is convex. To approximate $K_s(l_{non-grazing})/K_s(l_{grazing})$, we estimate the coupled attenuation coefficient $A$ in Eq. (1), once for grazing and once for non-grazing polar angles. Let the two estimates be $A_{grazing}$ and $A_{non-grazing}$ respectively. We then take the ratio $A_{grazing}/A_{non-grazing} \approx K_s(l_{non-grazing})/K_s(l_{grazing})$ as a measure of per-pixel convexity. Fig. 5 shows an example of the estimated convexity at each pixel of an input photometric stereo image. Though the exact height of each pixel is still unknown, the structures on the surface are clearly seen in this convexity map.
IV. RESOLUTION ENHANCEMENT

For resolution enhancement, our method divides the recovered distribution of normal directions among pixels at a higher resolution. For an $R \times R$ enhancement, with $M = R^2$, each higher resolution pixel covers $1/M$ of a pixel at the original resolution, such that $1/M$ of the recovered normal distribution should be assigned to each of the higher resolution pixels. We employ a simplified arrangement procedure where a mixture of $M$ uniform-weight Gaussians is estimated in the normal distribution recovery of Sec. III-A, and each component of the $M$-Gaussian GMM is assigned to one of the $M$ higher resolution pixels. The principal normal direction of each high resolution pixel is thus taken as the mean normal of the assigned GMM component.

To determine the spatial organization of these GMM components among the high resolution pixels, we employ constraints based on geometric consistency including integrability, pixel-wise convexity and minimum description length criteria. These constraints lead to a challenging optimization problem, which is solved using a combination of belief propagation and Markov Chain Monte Carlo (MCMC). To speed up convergence, the spatial arrangement of GMM components is first optimized by belief propagation with only integrability and pixel-wise convexity constraints. Then, the constraint from minimum description length criteria is added, and MCMC is applied to optimize the objective function.
A. Constraints on Normal Arrangement

Gaussian components should be arranged in a manner consistent with the convexity information computed in the preceding section. If the convexity value is small, we favor concave arrangements. Otherwise, convex arrangements are preferred. Another fundamental constraint in a normal map is that the curl of the normal map be equal to zero, which is known as the integrability constraint and is widely used in surface reconstruction (e.g., [25]–[27]).

These constraints do not provide sufficient information for determining a reliable solution, so we additionally take advantage of the observation that at a local scale there generally exists only a small number of perceptually distinct structural features on a surface. This surface property is the basis for work on 3D textons [12], which represent the appearance of points on a surface by indexing to a small vocabulary of prototype surface patches.

In our method, we utilize a constraint on normal arrangements that is motivated by the work on 3D textons and the minimum description length principle [11]. Specifically, our formulation favors normal arrangements that minimize the total number of local structural features, or 3D textons, needed to describe the imaged surface at the enhanced resolution. In [12], 3D texton primitives are represented in terms of co-occurring responses to a bank of Gaussian derivative filters of different orientations and spatial frequencies, but for our context of photometric stereo, we instead describe these structural primitives simply as a concatenated vector of principal normal orientations. In our implementation, the size of the texton area is set to be that of a pixel at the original, captured image resolution.

To minimize the number of distinct structural features in the surface description, we first cluster similar normal arrangements that may occur on the surface, and represent each cluster by a single representative arrangement, or texton. The set of GMM components for a given pixel has a number of possible arrangements, e.g., $4!$ arrangements for a $2 \times 2$ enhancement of resolution. As illustrated in Fig. 6, we take each possible arrangement for each low resolution pixel and plot it as a point in an arrangement space. These normal arrangements are grouped into texton clusters that are each represented by its mean vector. In clustering, each arrangement is represented by a concatenated vector of principal normal directions inside each pixel. For example, in the case of $2 \times 2$ resolution enhancement, a 8-D space is formed by concatenating four normalized principal normals each expressed as 2D vectors. In our implementation, the structure of this space is obtained by again employing the EM algorithm. We compute a Gaussian mixture model of these various arrangements.
Fig. 6. Mapping of possible normal arrangements of each pixel into a high dimensional arrangement space. Each possible arrangement of a pixel is a point in this space, indexed by concatenated normal vectors. Similar arrangements form clusters that represent possible textons of the surface.

Fig. 7. Texton equivalence classes. (a) Textons that differ in high-resolution structure but have a similar low-resolution distribution of normals are grouped into an equivalence class. (b) Each pixel belongs to one of the equivalent classes. A single texton is computed for each class, and the pixels of the class are assigned the geometric structure of this representative texton.
in the 8-D space. To distinguish it from the GMM model used for facet normal distribution, we refer this model as the texton GMM. In our algorithm, we fit a texton GMM with a large number of components and then eliminate components that have weights below a threshold, which is set to 0.01.

Distinct textons from this clustering may contain an identical set of principal normal directions that differs only in arrangement. This is exemplified in Fig. 6, where two different arrangements of the same normal distribution are grouped into different clusters. To minimize the geometric description, we group these textons into equivalence classes as shown in Fig. 7(a), and represent each class by a single texton. Each equivalence class contains a set of pixels that should be arranged according to its selected texton. These equivalence classes are determined by grouping texton clusters that are associated with the same pixels, since this indicates that the textons are rearrangements of each other. We compute this grouping with a voting scheme, in which the affinity of two textons is measured by the number of common pixels associated with the two sets of clusters. Thresholding this affinity measure gives a partitioning of textons into equivalence classes. With this partitioning, each pixel is initially associated with the equivalence class that contains the largest number of its possible normal arrangements, as illustrated in Fig. 7(b). When a texton is later assigned to represent an equivalence class, the pixels associated with the class will then be assigned the normal arrangement of that texton. Since formation of equivalence classes involves thresholding of an affinity measure, there potentially exist unreliable classifications of pixels to equivalence classes. As described in Section IV-B, this association between pixels and equivalence classes will be further refined in a later optimization procedure that involves local adjustments of individual low resolution pixels.

To determine the set of textons that are used to represent the set of equivalence classes, which we refer to as solution textons, we solve for the set of textons that best models the surface. Since any integrable texton from an equivalence class can accurately represent the pixels associated with the class, we determine the solution textons as those that are additionally consistent with the resulting structure in shifted pixels, which refer to pixel areas that overlap multiple pixels at low resolution as exemplified in Fig. 8(b). Based on our criterion for surface consistency, shifted pixels should also be represented by the solution textons. In the following section, we describe how a minimal texton description in conjunction with integrability constraints are used to compute the representative texton of each equivalence class.
B. Arrangement Optimization

To solve for the solution textons, and hence the arrangement of normals over the surface, we formulate objective functions that account for per-pixel convexity, texton description length, and integrability constraints.

As shown in Fig. 8(a), the convexity of the surface inside a closed curve \( S \) can be measured by the line integral \( \oint_N \cdot n ds \). Here, \( N \) is the normal field defined inside \( S \). If this integral is greater than zero, the surface is concave. Otherwise, it is convex. For each pixel \( \bar{s} = (x, y) \) and its associated texton \( \mathbb{L}(\bar{s}) \), we consider the curve that extends around a pixel as illustrated in Fig. 8(b), and compute the convexity as

\[
\text{Intg}_0(\bar{s}, \mathbb{L}(\bar{s})) = -a_x + a_y + b_x + b_y + c_x - c_y - d_x - d_y
\]

where \( a = (a_x, a_y, 1), b = (b_x, b_y, 1), c = (c_x, c_y, 1) \) and \( d = (d_x, d_y, 1) \) are the principal normal directions of the pixel. We compute this integration for all the possible textons of the pixel and normalize them linearly to \([-1, 1]\). An energy function with respect to the shadow-based convexity metric \( r = A_{\text{grazing}}/A_{\text{non-grazing}} \) can be defined as

\[
E_0(\bar{s}, \mathbb{L}(\bar{s})) = \begin{cases} 
\exp(-g_3(\hat{\text{Intg}}_0)) & \text{if } r > \lambda_2 \\
\exp(-g_2(\hat{\text{Intg}}_0)) & \text{if } \lambda_1 < r < \lambda_2 \\
\exp(-g_1(\hat{\text{Intg}}_0)) & \text{if } r < \lambda_1
\end{cases}
\]

Here, \( \lambda_1 < 0.7 < \lambda_2 \), and \( \hat{\text{Intg}}_0 \) is the normalized value of the integral \( \text{Intg}_0(\bar{s}, \mathbb{L}(\bar{s})) \). The profiles of functions \( g_1(\cdot), g_2(\cdot), g_3(\cdot) \) are illustrated in Fig. 9. When the convexity measure \( r \) is less than...
\[ \lambda_1, \text{the surface is concave, and } g_1(\cdot) \text{ is selected for the convexity energy term. Textons with a larger integration value are therefore favored. When } r \text{ is larger than } \lambda_2, \text{ the surface is convex, and textons with a small integration value are favored. If } r \text{ is close to 0.7, neither strongly concave nor strongly convex textons are wanted. Maximizing } E_0 \text{ will thus favor a normal arrangement consistent with the shadowing constraint. For all the examples, we use the settings } \lambda_1 = 0.3 \text{ and } \lambda_2 = 0.8. \]

Surface integrability is evaluated by the line integral \( \oint N \cdot t \, ds \), which should be zero for an integrable normal field. For a pixel \( \bar{s} = (x, y) \) and its texton \( \mathbb{L}(\bar{s}) \), this constraint can be expressed as

\[ \text{Intg}_1(\bar{s}, \mathbb{L}(\bar{s})) = a_x + b_x - b_y - d_y - d_x - c_x + c_y + a_y = 0. \]

An energy function for integrability can therefore be defined on each pixel as

\[ E_1(\bar{s}, \mathbb{L}(\bar{s})) = \exp \left( - \left( \text{Intg}_1(\bar{s}, \mathbb{L}(\bar{s})) \right)^2 \right). \]

Maximizing \( E_1 \) favors normal arrangements with minimal integration values.

For each pair of neighboring pixels \( \bar{s}, \bar{t} \) at the original resolution and their associated textons \( \mathbb{L}(\bar{s}), \mathbb{L}(\bar{t}) \), the integrability constraint also applies to the shifted pixel that overlaps \( \bar{s}, \bar{t} \), illustrated as a shaded region in Fig. 8(c). The integrability constraint in this instance can similarly be expressed as

\[ \text{Intg}_2(\bar{s}, \bar{t}, \mathbb{L}(\bar{s}), \mathbb{L}(\bar{t})) = e_x + f_x - f_y - h_y - h_x - g_x + g_y + e_y = 0. \]

A shifted pixel additionally should be associated to a solution texton. This condition can be quantified as \( \max_{1 \leq j \leq T} P_{ij}(\text{efgh}) \), where \( \text{efgh} \) is the concatenated vector of principal normals \( \text{e, f, g, h} \), \( \{ \mathbb{L}_{i1}, \mathbb{L}_{i2}, \ldots, \mathbb{L}_{iT} \} \) is the set of solution textons, and \( P_{ij}(\cdot) \) is the pdf function of the \( i_j \)-th Gaussian component of the texton GMM model. Then an energy function can be defined on
each pixel pair as
\[ E_2(\overline{s}, \overline{t}, \mathbb{L}(\overline{s}), \mathbb{L}(\overline{t})) = \exp \left( -\left( \text{Intg}_2(\overline{s}, \overline{t}, \mathbb{L}(\overline{s}), \mathbb{L}(\overline{t})) \right)^2 \right) \cdot \max_{1 \leq j \leq T} P_{ij}(\text{efgh}). \]
Maximizing \( E_2 \) will favor the selection of textons for \( \overline{s} \) and \( \overline{t} \) for which the normal arrangement of their shifted pixel is integrable and consistent with the set of solution textons.

In principle, integrability and an association to solution textons should also exist for other pixel displacements, e.g., a shifted pixel that overlaps the corners of four pixels. In our current implementation, these cases are not considered to avoid additional complexity in the energy formulation.

The solution textons \( \{\mathbb{L}_{i_1}, \mathbb{L}_{i_2}, \ldots, \mathbb{L}_{i_T}\} \) can be computed by maximizing the product of \( E_0, E_1 \) and \( E_2 \) over the whole image as
\[
\{\mathbb{L}_{i_1}, \mathbb{L}_{i_2}, \ldots, \mathbb{L}_{i_T}\} = \arg \max \prod_{\overline{s}} E_0(\overline{s}, \mathbb{L}(\overline{s})) E_1(\overline{s}, \mathbb{L}(\overline{s})) \prod_{Nb(\overline{s}, \overline{t})} E_2(\overline{s}, \overline{t}, \mathbb{L}(\overline{s}), \mathbb{L}(\overline{t}))
\]
where \( \overline{s} \) indexes all the pixels in the image, and \( Nb(\overline{s}, \overline{t}) \) represents all pairs of 4-neighbor pixels. Due to the complexity of this optimization problem, a solution is obtained using a two-step process. The second step uses a Markov Chain Monte Carlo method to compute an arrangement solution that accounts for both texton and integrability constraints. To aid MCMC in reaching a good solution, the first step formulates the shadowing and integrability constraints in a graphical network and employs belief propagation to compute a good initial solution for input into the MCMC process.

1) Initialization with Belief Propagation: In this first step, the problem is formulated as an undirected graph model, where each node \( v \) of the graph represents an equivalence class. Two nodes are connected if and only if they contain pixels that are 4-neighbors of each other in the image. Each node has a number of candidate textons that are indexed by labels. Integrability constraints are applied to the graphical model as energy functions defined on nodes and edges.

From the preceding discussion, we can define an energy term for each label \( \mathbb{L} \) of a node as
\[
E_{\text{node}}(\mathbb{L}; v) = \prod_{\overline{s} \in v} E_0(\overline{s}, \mathbb{L}) E_1(\overline{s}, \mathbb{L})
\]
where \( \overline{s} \in v \) denotes pixels \( \overline{s} \) in equivalence class \( v \). For each pair of connected nodes \( v_1, v_2 \) and their labels \( \mathbb{L}_{1}, \mathbb{L}_{2} \) in the graphical model, we define an energy on the connecting edge as
\[
E_{\text{edge}}(\mathbb{L}_{1}, \mathbb{L}_{2}; v_1, v_2) = \prod_{Nb(\overline{s}, \overline{t}); \overline{s} \in v_1; \overline{t} \in v_2} E_2(\overline{s}, \overline{t}, \mathbb{L}(\overline{s}), \mathbb{L}(\overline{t}))
\]
where \( E'_2 \) denotes the energy \( E_2 \) without the texton constraint \( \max_{1 \leq j \leq T} P_i, (efgh) \). The initial solution is determined by maximizing the energy over the entire graphical model using the belief propagation algorithm in [14].

2) MCMC Optimization: The second step takes the solution of the first step as an initialization to an MCMC process with the energy function

\[
E_{\text{total}} = \prod \bar{s} E_0(\bar{s}, \mathbb{L}(\bar{s})) E_1(\bar{s}, \mathbb{L}(\bar{s})) \prod_{N(\bar{s}, \bar{t})} E_2(\bar{s}, \bar{t}, \mathbb{L}(\bar{s}), \mathbb{L}(\bar{t})).
\]

Given a function \( \pi(x) \), in our case \( \pi(x) = E_{\text{total}} \), MCMC is a strategy for generating samples \( \{x^k\}_{k=1}^K \) to explore the state space of \( x \) using a Markov chain mechanism. More samples are generated in areas where \( \pi(x) \) is larger. The stationary distribution of the chain is \( \pi(x) \). The maximum value of \( \pi(x) \) can therefore be found as

\[
x^* = \arg \max \pi(x^k).
\]

The efficiency of MCMC depends on the proposal function \( q(x^*|x^k) \), which gives the probability of transferring from state \( x^k \) to \( x^* \).

We construct a mixture of two kinds of proposals: a local proposal \( q_l \) and a global proposal \( q_w \). The global proposal explores vast regions of the state space, and local proposals search among finer details of the energy function. The final proposal is defined as

\[
q(x^*|x^k) = \alpha_l q_l(x^*|x^k) + \alpha_w q_w(x^*|x^k)
\]

where \( \alpha_l + \alpha_w = 1 \) are positive weights. Both local and global proposals are based on the Metropolized Independence Sampler (MIS) [15].

The local proposal changes the equivalence class association of one low resolution pixel at a time. A low resolution pixel is originally associated with the equivalence class that contains the largest number of its possible normal arrangements. In practice, this association could be problematic due to factors such as imaging noise and thresholding in equivalence class formation. To fine-tune the arrangement, we allow each pixel to change its equivalence class with a certain probability. The local proposal randomly selects a low resolution pixel, changes its associated equivalence class according to a given probability, and uses the texton of the new equivalence class as the arrangement for the pixel. First, a pixel is selected according to the energy of its current texton. The energy of a pixel \( \bar{s} \) with texton \( \mathbb{L}_1 \) is

\[
E_{\bar{s}}(\mathbb{L}_1) = E_0(\bar{s}, \mathbb{L}_1) E_1(\bar{s}, \mathbb{L}_1) \prod_{N(\bar{s}, \bar{t})} E_2(\bar{s}, \bar{t}, \mathbb{L}(\bar{s}), \mathbb{L}(\bar{t})).
\]

The probability \( w_{\bar{s}} \)

DRAFT
for it to be selected is defined as

\[ w_s = \frac{E_s(L_1)}{\sum_i E_i(L(t))}. \]

A candidate texton \( L_2 \) in the current solution texton set is proposed with a probability defined by the Gaussian distribution \( g_s(L_2) = g(E_s(L_2); \mu_{E_s}, \sigma_{E_s}^2) \). Here, \( \mu_{E_s} \) is the minimum energy among all the possible textons of \( \bar{s} \), \( \sigma_{E_s} \) is set to the difference between \( \mu_{E_s} \) and the median energy among the textons. This proposal is accepted with probability

\[ \frac{E_s(L_2)g_s(L_2)}{E_s(L_1)g_s(L_1)}. \]

To give the Markov chain the ability to jump out of local minima, we design a global proposal that simultaneously changes the arrangement of all the low resolution pixels associated with a given equivalence class. Here, an equivalence class is randomly selected and its texton is changed randomly. Similar to the local proposal, an equivalence class is first selected according to its current texton setting. The energy induced by an equivalence class \( S \) with its current configuration \( L_1 \) is \( E_S(L_1) = \prod_{\bar{s} \in S} E_s(L_1) \). The equivalence class is selected with probability

\[ w_s = \frac{E_S(L_1)}{\sum_{T} E_T(L(T))}. \]

Similarly, a candidate \( L_2 \) is drawn with probability \( g_S(L_2) = \prod_{\bar{s} \in S} g_{s}(L_2) \) and accepted with probability

\[ \frac{E_S(L_2)g_S(L_2)}{E_S(L_1)g_S(L_1)}. \]

The relative weights of the local and global control coefficients \( \alpha_l, \alpha_w \) represent the frequencies of local and global proposals. Generally, a global proposal should be made only after numerous local proposals fail to improve the result. In our implementation, we set \( \alpha_l = \exp\left(-\frac{\text{iter}^2}{2\sigma^2}\right) \), where \( \text{iter} \) denotes the number of iterations over which local proposals have not improved the result. \( \sigma \) controls the probability that a local proposal is utilized, and is set to 100.

With this approach, various degrees of resolution enhancement can be obtained, but greater amounts of enhancement lead to substantial increases in computation, due to the \( R^{2!} \) possible normal arrangements for an \( R \times R \) enhancement. To alleviate this problem, an approximation can be employed where \( 2 \times 2 \) enhancements are performed iteratively to reach higher levels of resolution.
Fig. 10. Example of a synthesized surface, with increasingly finer surface structure towards the center. Images are normalized in size to facilitate comparison. To view finer detail, please zoom in on the electronic version. (a) One of the photometric stereo images; (b) Height field recovered at the original resolution; (c) Height field recovered at a $2 \times 2$ enhanced resolution; (d) Height field recovered at a $4 \times 4$ enhanced resolution.

Fig. 11. Example of a stone carving. Images are normalized in size to facilitate comparison. To view finer detail, please zoom in on the electronic version. (a) One of the photometric stereo images; (b) Height field recovered at the original resolution; (c) Height field recovered at a $2 \times 2$ enhanced resolution; (d) Height field recovered at a $4 \times 4$ enhanced resolution.

Fig. 12. Example of a shiny metal surface. Images are normalized in size to facilitate comparison. To view finer detail, please zoom in on the electronic version. (a) One of the photometric stereo images; (b) Height field recovered at the original resolution; (c) Height field recovered at a $2 \times 2$ enhanced resolution; (d) Height field recovered at a $4 \times 4$ enhanced resolution.
Fig. 13. Example of a wood flower frame with shiny paint. Images are normalized in size to facilitate comparison. To view finer detail, please zoom in on the electronic version. (a) One of the photometric stereo images; (b) Height field recovered at the original resolution; (c) Height field recovered at a $2 \times 2$ enhanced resolution; (d) Height field recovered at a $4 \times 4$ enhanced resolution.

Fig. 14. Quantitative analysis of reconstruction accuracy at different enhancement levels on synthesized data. The ground truth height map is $\cos(40\theta)/\rho$, where $\theta, \rho$ are polar coordinates. The frequency of geometry fluctuations, i.e., the resolution of subpixel geometric structure, gradually increases toward the origin. We plot the angular error with respect to distance from the origin, which reveals the reconstruction accuracy at different resolutions. The chart shows error-distance curves for reconstruction with the original resolution, $2 \times 2$ enhancement, and $4 \times 4$ enhancement.
Fig. 15. Renderings of the recovered height fields from a frontal viewpoint. Left to right: original resolution, $2 \times 2$ enhancement, $4 \times 4$ enhancement, and ground truth computed from a close-up image sequence approximately equivalent to a $3 \times 3$ enhancement level.
V. RESULTS

In our experiments, images of a surface are captured by a fixed camera with illumination from seven incandescent bulbs attached to an arc that rotates around the surface. The seven bulbs are placed at equal angular intervals on the arc, and the arc rotates at intervals of 30 degrees. At two of the arc rotation angles, lighting is occluded by the capture device, and other occlusions of illumination occasionally occur. Generally, 60–70 photometric stereo images are captured for each surface with our device.

We applied our method to one synthesized example in Fig. 10 and a few real surfaces, displayed in Fig. 11, Fig. 12, and Fig. 13. For each of the figures, an input image at the captured resolution is shown. For visualization purposes, normal maps are integrated into height fields according to the method in [26]. Alternative techniques for height field construction from normal maps may be used instead [28], [29]. Height fields computed at the captured resolution exhibit little geometric detail. With $2 \times 2$ resolution enhancement, sub-pixel structures lost in conventional photometric stereo
are recovered. $4 \times 4$ enhancement adds further detail. As greater enhancement is applied, some increase in noise is evident, partially due to the greater complexity in determining proper normal arrangements. Noise may also arise from deviations of the actual reflectance from the mathematical model we used for normal recovery.

Fig. 14 provides a quantitative analysis on results for the synthesized data. The underlying ground truth height map of the surface is $\cos(40\theta)/\rho$, where $\rho, \theta$ are polar coordinates. The frequency of geometry fluctuations, i.e., the resolution of subpixel geometric structure, gradually increases toward the origin. The average angular error between the reconstructed normal and ground truth normal is plotted as a function of the distance to the origin. This error increases towards the origin because of the greater complexity in subpixel geometric structure. As shown in Fig. 14, our method can significantly improve reconstruction accuracy in comparison to a conventional photometric stereo method based on our reflectance model. However, at a very high resolution of subpixel structure, i.e., close to the origin of the example surface, our $2 \times 2$ enhancement does not outperform conventional photometric stereo, because each subpixel normal does not adequately represent its underlying structure structure. The result of $4 \times 4$ enhancement can represent more complicated subpixel structure and further improve reconstruction accuracy, but due to the complexity of arranging 16 normals, this improvement is less significant. The numerous possible arrangements also complicates reconstruction for regions distant from the origin for $4 \times 4$ enhancement.

We also examined the reconstruction accuracy in Table I for the real surface shown in Fig. 11. For evaluation purposes, we downsample the original photometric stereo images for processing by our resolution enhancement algorithm, and compare our results to ground truth obtained by conventional photometric stereo on the original images. The average reconstruction error is reduced with our resolution enhancement, especially with the convexity constraint, though the quantitative improvement over conventional photometric stereo on the downsampled images is not as dramatic as what can be seen visually. This is due to the presence of many pixels without much subpixel structure, for which our method and conventional photometric stereo performs similarly. Enhancement at the $4 \times 4$ level results in greater reconstruction error than with $2 \times 2$ enhancement because of arrangement complexity. Visually, $4 \times 4$ enhancement provides a sharper height map, but also contains more noise from arrangement errors.

A different visualization of the results is provided in Fig. 15 and Fig. 16. We render the recovered height fields from a frontal and a slanted viewpoint for the original resolution, $2 \times 2$ enhancement,
<table>
<thead>
<tr>
<th>Resolution</th>
<th>Method</th>
<th>Average Angular Error (degree)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Downsized Resolution</td>
<td>Conventional P.S.</td>
<td>9.2</td>
</tr>
<tr>
<td>$2 \times 2$ Enhancement</td>
<td>Without Convexity</td>
<td>8.8</td>
</tr>
<tr>
<td>$2 \times 2$ Enhancement</td>
<td>With Convexity</td>
<td>8.2</td>
</tr>
<tr>
<td>$4 \times 4$ Enhancement</td>
<td>Without Convexity</td>
<td>8.9</td>
</tr>
<tr>
<td>$4 \times 4$ Enhancement</td>
<td>With Convexity</td>
<td>8.4</td>
</tr>
</tbody>
</table>

TABLE I
COMPARISON OF RECONSTRUCTION ACCURACY FOR THE REAL SURFACE SHOWN IN FIG. 11.

$4 \times 4$ enhancement, and the ground truth computed from a close-up image sequence approximately equivalent to a $3 \times 3$ enhancement level. Geometric detail that is seen in the ground truth height field becomes increasingly clearer with greater enhancement.

VI. CONCLUSION

We proposed a method to enhance the resolution of photometric stereo by recovering a general normal distribution and the surface convexity of each pixel, and then arranging these normals spatially within the pixel according to consistency and simplicity constraints on surface structure. With this approach, fine-scale surface structure that is missing in conventional photometric stereo can be inferred from low-resolution input.

In future work, we plan to examine directions for potentially enhancing the quality of our subpixel reconstruction results. One is to enforce greater consistency of geometric structure over a surface by optimizing normal arrangements in a manner that more comprehensively accounts for integrability and texton constraints, such as over low resolution pixel areas with arbitrary subpixel shifts. Another is to examine methods for utilizing additional input data, such as from subpixel viewpoint shifts, in conjunction with our subpixel photometric stereo approach.

ACKNOWLEDGMENT

The work is supported by Hong Kong RGC projects 619005 and 619006.

REFERENCES

Ping Tan received the B.S. degree in Applied Mathematics from Shanghai Jiao Tong University, China, in 2000. He received the Ph.D. degree in Computer Science and Engineering from the Hong Kong University of Science and Technology in 2007. His research interests include image-based modeling, photometric stereo and reflectance modeling.

Stephen Lin received the B.S.E degree in electrical engineering from Princeton University, Princeton, NJ, and the Ph.D. degree in computer science and engineering from the University of Michigan, Ann Arbor.

He is a Lead Researcher in the Internet Graphics Group, Microsoft Research Asia. His current research interests include photometric analysis in computer vision and reflectance modeling in computer graphics.
Long Quan received the Ph.D. degree in Computer Science from INPL-CNRS-INRIA, France, in 1989. Before joining the Department of Computer Science at the Hong Kong University of Science and Technology (HKUST) in 2001, he has been a CNRS senior research scientist at INRIA. His research interests include vision geometry, 3D reconstruction, structure from motion, and image-based modeling and rendering. He is a regional editor of Image and Vision Computing Journal (IVC), and is on the editorial board of the International Journal of Computer Vision (IJCV), Machine Vision and Applications (MV A), the Electronic Letters on Computer Vision and Image Analysis (ELCVIA), and Foundations and Trends in Computer Graphics and Vision. He has served as an associate editor of the IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI); as a programme co-chair of IAPR International Conference on Pattern Recognition 2006 (ICPR) Computer Vision and Image Analysis, and as an area chair of International Conference on Computer Vision (ICCV), European Conference on Computer Vision (ECCV), and IEEE Computer Vision and Pattern Recognition (CVPR).