Shadow Graphs and 3D Texture Reconstruction

YIZHOU YU

Department of Computer Science, University of Illinois at Urbana-Champaign, 201 North Goodwin Avenue, Urbana, IL 61801, USA

yyz@cs.uiuc.edu

JOHNNY T. CHANG*

Jet Propulsion Laboratory, Mailstop 198-235, 4800 Oak Grove Drive, Pasadena, California 91109, USA

johnny.t.chang@jpl.nasa.gov

Abstract. We present methods for recovering surface height fields such as geometric details of 3D textures by incorporating shadow constraints. We introduce shadow graphs which give a new graph-based representation for shadow constraints. It can be shown that the shadow graph alone is sufficient to solve the shape-from-shadow problem from a dense set of images. Shadow graphs provide a simpler and more systematic approach to represent and integrate shadow constraints from multiple images. To recover height fields from a sparse set of images, we propose a method for integrated shadow and shading constraints. Previous shape-from-shadow algorithms do not consider shading constraints while shape-from-shading usually assumes there is no shadow. Our method is based on collecting a set of images from a fixed viewpoint as a known light source changes its position. It first builds a shadow graph from shadow constraints from which an upper bound for each pixel can be derived if the height values of a small number of pixels are initialized correctly. Finally, a constrained optimization procedure is designed to make the results from shape-from-shading consistent with the height bounds derived from the shadow constraints. Our technique is demonstrated on both synthetic and real imagery.

Keywords: 3D texture, surface geometry, shape-from-shadow, shadow graph, shading, photometric stereo, optimization

1. Introduction

The visual appearance of an object is governed by a hierarchy of geometric components, including a *macrostructure* layer specified as a set of polygonal and/or curved surfaces, a *mesostructure* layer involving geometric details that are relatively small but still individually visible such as bumps and dents on a surface, and a *microstructure* layer that involves surface

microfacets visually indistinguishable from each other. Traditional surface appearance models include bidirectional reflectance distribution functions (BRDFs) at the microstructure level, and bump maps at the mesostructure level. Combined with geometric models of macrostructure, they can be used to produce high-quality renderings, and provide novel visual interactions between humans and machines.

By *3D textures*, we mean small-scale surface details that include both the mesostructure layer and the microstructure layer. The microstructure layer leads to pointwise surface reflectance properties while the

mesostructure layer gives rise to mutual shadowing and occlusions as well as pointwise local surface normal orientations. Both layers contribute to the color variations on a macrostructure surface. A photograph of the color variations on a surface is a conventional (2D) texture defined by the computer vision and image processing communities. Even with a fixed 3D texture on a surface, this conventional texture can change with respect to different lighting and viewing conditions. The mesostructure layer of a 3D texture can often be modeled as a height field. From another point of view, a 3D texture is nothing but a height field with a distinct reflectance function associated with every point. If the height field is Lambertian, pointwise reflectance functions degenerate into pointwise albedos.

Among the aforementioned three scales of geometry, modeling the mesostructure layer has caught researchers' attention recently along with other image-based shape and texture modeling problems (Van Gool et al., 2002). The visual effects caused by 3D textures lead to object appearance changes under different lighting and viewing directions. Traditional texture mapping or bump mapping techniques in computer graphics fall short at reproducing the mutual shadowing and occlusion effects. Recently, bidirectional texture functions (BTFs) (Dana et al., 1999) have been proposed to capture and reproduce these visual effects. Actually 3D textures and BTFs are closely related. A 3D texture exhibits inherent geometric and photometric properties while a BTF represents the projections of a 3D texture in the 2D image plane under all possible illumination and viewing directions. We can always obtain a BTF from a 3D texture by applying graphics rendering processes. While a BTF is a six-dimensional function which requires a large amount of data storage, a 3D texture can also be represented discretely as a 2D map of height displacements along with reflectances. The latter representation offers a better choice as the internal representation for 3D textures. However, height displacements and reflectances for real 3D textures are not directly available without applying reconstruction techniques from computer vision. The methods for recovering height displacements have so far been limited to photometric stereo and shape-from-shading. Surface albedos can be obtained as by-products during these recovery processes.

In this paper, we consider 3D texture surface reconstruction by incorporating shadows. Shape-from-shadow tries to reconstruct a surface using multiple shadow images. It has a few advantages compared to other surface reconstruction techniques. For example, shadow constraints are insensitive to specular reflection and spatial variations of reflectance, and are able to impose long-range height constraints. The basic conclusion from previous work along this direction (Medioni, 1983; Shafer, 1985; Kender and Smith, 1987; Hatzitheodorou, 1989; Daum and Dudek, 1998; Kriegman and Belhumeur, 1998) says that with a sufficient number of shadow images, the underlying surface can be recovered. However, the proposed algorithms for this problem are either complicated or heuristic. The major reason for this is that it was not clear how to effectively represent shadow constraints and integrate the information from multiple shadow images.

To clearly understand this problem, we introduce shadow graphs (Yu and Chang, 2002) which can effectively represent and integrate shadow constraints from multiple images. We prove that the shadow graph alone is sufficient to solve the shape-from-shadow problem from a dense set of images. Simple operations on a shadow graph enable us to derive the structures of the underlying surface. This approach is simpler and more systematic than the previous methods.

Usually most of the pixels in an image are not shadowed. However, shape-from-shadow neglects the rest of the shading information in the input images. As we will see, shadow constraints are usually inequalities which are not as powerful as equalities. Consequently, it usually requires a dense set of input images to obtain good results. On the other hand, shape-from-shading (Horn and Brooks, 1986) and photometric stereo (Woodham, 1989) are effective approaches for a large class of surfaces including faces and sculptures. Both techniques use the pixelwise shading information to constrain surface normals, and do not allow shadows in the input images. They need an integration step to reconstruct a surface. This step tends to accumulate errors from pixel to pixel. Although theoretically they can uniquely recover the underlying surface, the final relative height differences between distant points may not come out very accurately.

To take the advantages from both shape-from-shadow and shape-from-shading, we also develop a method of recovering shape from both shadow and shading constraints. A constrained optimization procedure is developed to make the results from shape-from-shading consistent with the height bounds derived from shadow constraints.

The rest of the paper is organized as follows. Related work on surface reconstruction using photometric

methods will be discussed in the next section. In Section 2, we introduce shadow graphs and present both theoretical and experimental results on height field reconstruction using shadow graphs only. In Section 3, we will consider integrating both shadow and shading constraints, and propose a hybrid method for surface reconstruction. Extensive experimental results on 3D textures and other surfaces using the hybrid method will also be reported. We will conclude this paper in Section 4.

1.1.   Related Work

One method to study 3D textures is through BTFs which can be regarded as a mapping from the 4D space of lighting and viewing directions to the space of all 2D images. Previous work on BTFs aims to capture appearance data for natural materials and represent them efficiently (Dana et al., 1999; Dana and Nayar, 1998, 1999; Leung and Malik, 1999; Dong and Chantler, 2002). Each material in the CUReT database (Dana et al., 1999) has a 4D sampling of images under various viewing and lighting conditions. Because of the changing viewing direction, the images in the CUReT database are not registered. The PhoTex database reported in Dong and Chantler (2002) has images of material samples taken from a fixed camera position. Therefore, it is more suitable for photometric stereo and surface reconstruction in general. Although the work in Liu et al. (2001) focuses on BTF synthesis, the first step of the process is to recover a surface height field for the considered material sample using shape-from-shading. A variety of options using intensities, normals and height values have been proposed in Dong and Chantler (2002) for high-quality 3D texture synthesis.

In terms of surface reconstruction, a few algorithms explicitly make use of shadow constraints (Kender and Smith, 1987; Hatzitheodorou, 1989; Daum and Dudek, 1998; Stewart and Langer, 1997). Most of them belong to shape-from-shadow(darkness) algorithms. Some shape-from-shadow algorithms (Kender and Smith, 1987) use a shadowgram as an intermediate representation which is derived from a dense set of lighting directions. Hatzitheodorou (1989) assumes the underlying surface has a spline representation because shadows only provide a relatively sparse set of constraints. The number of unknown coefficients in the spline model is designed to scale with the number of shadow constraints. Daum and Dudek (1998) introduces a shape-from-shadow algorithm us-

ing relaxation. A pair of upper-bound and lower-bound surfaces are constructed by updating the height values at pixels with violated shadow constraints. Like shape-from-shading, shape-from-shadow can recover unknown lighting directions as well (Kriegman and Belhumeur, 1998).

The computation of shape-from-shading has been typically characterized as that of finding surface orientation from one single image followed by a step that converts the orientation information into height under integrability constraints. The surface is usually assumed to be Lambertian. Leclerc and Bobick (1991) introduces an algorithm that allows direct computation of height from shading. Since the unknowns directly represent pixelwise height values, this approach can be more naturally integrated with other methods of recovering shape, such as stereo and shape-from-shadow. Dupuis and Oliensis (1994) presents provably convergent algorithms for this problem.

Photometric stereo (Woodham, 1989) can usually obtain better results than shape-from-shading because of the use of multiple input images. This approach has been generalized to recover shape for metallic and hybrid surfaces with both diffuse and specular reflection (Ikeuchi, 1981; Nayar et al., 1990). The lighting direction for each image is usually assumed to be known. However, both surface shape and lighting directions can be recovered simultaneously from SVD decomposition up to a bas-relief transformation (Belhumeur et al., 1997; Belhumeur and Kriegman, 1998). Shadowed pixels in each image can be masked out in the process with the hope that there are still enough images covering them (Jacobs, 1997; Georghiades et al., 1999).

Stewart and Langer (1997) considers recovery of shape from shading under a uniform hemispherical light source. Partial shadowing is taken into account because only a part of the light source is visible from every surface point. Interreflections are also considered in the algorithm presented in Nayar et al. (1991).

2.   Shadow Graphs

We consider recovering terrain-like surface height fields for 3D textures in this paper. For the convenience of discrete representation based on pixels, a height field is assumed to be a piecewise constant function with every pixel corresponding to a piece with constant height. Every piece of the height field is represented by the point corresponding to the center of the pixel. We assume the incident lighting direction from a collimated

source is known. This assumption is reasonable because the lighting direction is usually carefully calibrated before 3D texture samples are captured. A calibrated lighting direction tends to be more accurate than a recovered one. We also assume that the distance between the camera and the surface is sufficiently large so that the orthographic projection model is accurate.

Let us first check what kind of constraints are available from images with shadows. Let $h(x)$ be a height field defined on a planar domain D with a finite area in the image plane and $\mathbf{L}$ be the lighting direction pointing downwards with a zenith angle (the angle of the lighting direction with respect to the global surface normal) $\theta(< 90°)$. The normal orientation of this height field is denoted as $\mathbf{n}(x)$. The boundary curve of domain $\mathbf{D}$ is Ω. The projected vector of $\mathbf{L}$ in the domain D is $\mathbf{L}_p$. Let $\mathbf{x}_i$ and $\mathbf{x}_j$ be two arbitrary 2D points in D. The line segment between them is denoted as a vector interval $[\mathbf{x_i}, \mathbf{x_j}]$ for convenience. Based on whether a point on the height field is in shadow or not under lighting direction $\mathbf{L}$, there are two different sets of constraints (Fig. 1).

– If any point on the line segment $[\mathbf{x_0}, \mathbf{x_1}]$ is in shadow, the points at $\mathbf{x}_0$ and $\mathbf{x}_1$ are the delimiting points of this shadow segment, and the point at $\mathbf{x}_0$ is the occluding point generating this shadow segment, we have the following *shadow* constraints.

$$h(\mathbf{x}) \leq h(\mathbf{x_0}) - \frac{\|\mathbf{x} - \mathbf{x_0}\|_2}{\tan\theta}, \quad \forall \mathbf{x} \in [\mathbf{x_0}, \mathbf{x_1}]; \quad (1)$$

$$h(\mathbf{x_1}) = h(\mathbf{x_0}) - \frac{\|\mathbf{x_1} - \mathbf{x_0}\|_2}{\tan\theta}; \quad (2)$$

$$\mathbf{L} \cdot \mathbf{n}(\mathbf{x_0}) = 0 \quad (3)$$

where the last equation means the lighting vector $\mathbf{L}$ falls inside the tangent plane at $\mathbf{x}_0$ if the original continuous height field is locally differentiable at $\mathbf{x}_0$.

– If the point at $\mathbf{x}_2$ is not in shadow, we have the following *antishadow* constraints.

$$h(\mathbf{x}) \leq h(\mathbf{x_2}) + \frac{\|\mathbf{x} - \mathbf{x_2}\|_2}{\tan\theta}, \quad \forall \mathbf{x} \in [\mathbf{x_b}, \mathbf{x_2}] \quad (4)$$

where $\mathbf{x_b} \in \Omega$ and the line segment $[\mathbf{x_b}, \mathbf{x_2}]$ is in the same direction as $\mathbf{L}_p$.

Let us first focus on how to represent the inequality constraints (1) and (4) in a graph.

Definition 1. A *shadow graph* is a weighted directed graph $G = (V, E, W)$ where the set of nodes V is the set
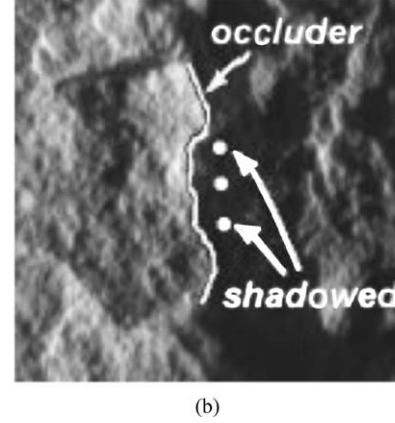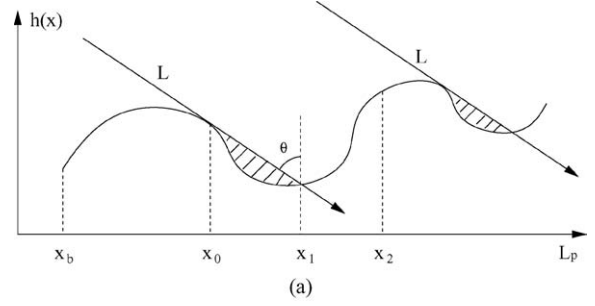

(a)

(b)

Figure 1. (a) 2D schematic of shadowed and nonshadowed regions on a terrain-like surface. L is the parallel lighting direction. x_0 is an occluder, x_1 is on the shadow boundary caused by x_0, and x_2 is a non-shadowed point. (b) An occluding contour and shadowed points in a real image.

of points defined on domain D, an edge $e = (\mathbf{x_i}, \mathbf{x_j}) \in E$ indicates $h(\mathbf{x}_j)$ is dependent on $h(\mathbf{x}_i)$ and $h(\mathbf{x}_i) - h(\mathbf{x}_j) \geq W(e)$ where the edge weight $W(e)$ can be any real number.

A shadow graph can be induced from an image of the height field under an arbitrary lighting direction $\mathbf{L}$. Shadowed pixels can be detected from the image, and an occluder can be located for each continuous shadow segment with the knowledge of the lighting direction. For example, if $[\mathbf{x_i}, \mathbf{x_j}]$ is a shadow segment and the vector from $\mathbf{x}_i$ to $\mathbf{x}_j$ is in the direction of the projected lighting direction $\mathbf{L}_p$, the point at $\mathbf{x}_i$ is the occluder of all the points in $[\mathbf{x_i}, \mathbf{x_j}]$. There should be an edge $(\mathbf{x_i}, \mathbf{x})$ with weight $\frac{\|\mathbf{x} - \mathbf{x_i}\|_2}{\tan\theta}$ in the induced graph for all $\mathbf{x} \in (\mathbf{x_i}, \mathbf{x_j}]$. This graph basically encodes the shadow constraints available from the image. All the edge weights in this graph should be positive. However, this graph can have negative weights if the additional antishadow constraints in Eq. (4) are represented as well.

Suppose we have multiple images of the height field under a set of lighting directions $\{\mathbf{L}_k\}_{k=1}^m$. Each of the images has its own shadow graph. Finally, the edges from all of these individual graphs can be accumulated into one graph that is corresponding to all the images. Since all the individual graphs have the same set of nodes, the merged graph simply has all the weighted edges from all of the individual graphs. Note that this merged graph does not have the specific lighting information, which is not particularly important because all the constraints essential to the height field are kept there.

Proposition 1.   *A shadow graph with positive weights is a directed acyclic graph.*

Proof: We can prove this by contradiction. Suppose there is a circular path in the graph and a node v is on the path. Since all the arcs on this path have positive weights, it is easy to conclude that $h(v) > h(v)$ by starting from v, going through this path and back to v. This is a contradiction.   □

Definition 2.   The transitive closure of a shadow graph G is defined to be a new graph $G^c = (V, E^c, W^c)$ on the same set of nodes such that $(\mathbf{x_i}, \mathbf{x_j}) \in E^c$ as long as there is a path from $\mathbf{x_i}$ to $\mathbf{x_j}$ in G, and $W((\mathbf{x_i}, \mathbf{x_j}))$ is set to be the maximum accumulated weight among the paths from $\mathbf{x}_i$ to $\mathbf{x}_j$.

There are a set of nodes $V_t \subset V$ in G^c that do not have any incident edges with positive weights, which means they are not shadowed by any other points in any of the images. The highest point(s) of the height field surely belongs to this set because there are no other points which can occlude it(them) from the light sources. The absolute height values of the nodes in V_t are unrecoverable from shadow constraints. However, if we can recover their height values from other approaches such as stereo processing, the information embedded in G^c can be used for obtaining an upper bound of the height at any point in $V - V_t$. The set of edges in G^c connecting V_t and $V - V_t$ becomes the most important for this purpose. Suppose there is a node $v \in V - V_t$ and a set of associated edges $E_v \subset E^c$ such that if an edge $e = (v_t, v) \in E_v$, $v_t \in V_t$. The upper bound of the height at the point corresponding to node v can be obtained from

$$U(h(v)) = \min_{(v_t,v)\in E_v}(h(v_t) - W((v_t, v))). \quad (5)$$

Let us examine the asymptotic behavior of this upper bound when we increase the number of input images with lighting directions covering the whole lighting hemisphere. The set V_t will shrink and approach its limit which is the set of the highest points of the height field. Otherwise, assume there is a pair of nodes $v_1, v_2 \in V_t$ and $h(v_2) < h(v_1)$. We can always design a lighting direction from which the point corresponding to v_1 shadows the point corresponding to v_2, which means $v_2 \notin V_t$. This is a contradiction. Since eventually V_t only has nodes at the same height, we do not need to seek their relative height through other reconstruction techniques. Our interest should be focused on the relative height of other points compared to the highest points whose height can always be set to zero.

Proposition 2.   *Equation* (5) *gives an upper bound for the height at any node in $V - V_t$ provided that the estimation of the height for the nodes in V_t is accurate. With an increasing number of input images with lighting directions covering the whole lighting hemisphere, Eq.* (5) *converges asymptotically to the correct relative height, with respect to the highest points in V_t, at any point in $V - V_t$.*

Proof: The first part is obvious. The second part can be proved by induction. Since we only have a finite number of points according to our surface model, we can sort the points in decreasing order of their height. The highest points in the sorted list are assumed to be at height zero. Suppose the point at $\mathbf{x}_m$ is the k-th element in the sorted list and the height of its $k - 1$ preceding elements can be recovered to an arbitrary precision independently of the height of the rest of the elements in the list. Now we show that the height of the point at $\mathbf{x}_m$ can also be recovered to an arbitrary precision independently of the height of its following elements in the list. Note that all the surface points are lit if we have a vertical lighting direction. If we decrease the angle of elevation of the light, the point at $\mathbf{x}_m$ will certainly be shadowed since it is not one of the highest points. Given a certain sampling density for the illumination hemisphere which contains all the possible lighting directions, there exist two adjacent sampled directions $\mathbf{L}_r$ and $\mathbf{L}_l$ such that the point at $\mathbf{x}_m$ is non-shadowed when the light is at $\mathbf{L}_r$ and becomes shadowed when the light moves to $\mathbf{L}_l$. An upper bound for this point can be obtained from $\mathbf{L}_l$ and an occluder at $\mathbf{x}_o$ whose height is recovered to an arbitrary precision. If we keep increasing the sampling density on the illumination hemisphere,

the difference between $\mathbf{L}_r$ and $\mathbf{L}_l$ can become arbitrarily small and the upper bound for the point at $\mathbf{x}_m$ can also become arbitrarily close to its true height.   □

The conclusion drawn from the above proposition is that the height field can be recovered to a prescribed precision from the shadow graph alone as long as we have sufficient number of images that evenly cover the illumination hemisphere above the material sample. Since the convergence can only be achieved asymptotically, a very large number of images would be required for a reasonable reconstruction.

It is clear that the antishadow constraints can be derived from the shadow constraints if we have a very dense set of images since the height field itself can be recovered from the shadow constraints alone according to the above Proposition. However, if we only have a sparse set of images, this is not necessarily true. Representing these antishadow constraints in a shadow graph usually can provide additional information. According to Eq. (4), antishadow constraints transform to additional edges with negative weights. Cycles can appear in the resulting graph. However, the accumulated weight of any cycle can not be positive according to the following Proposition.

Proposition 3. *The accumulated weight of a circular path in a shadow graph must be either zero or negative.*

Proof: Suppose $\mathbf{x_0}, \mathbf{x_1}, \ldots, \mathbf{x_n} \in V$ are consecutive nodes of a circular path, i.e. $(\mathbf{x_i}, \mathbf{x_{i+1}}) \in E(i = 0, \ldots, n-1)$ and $(\mathbf{x_n}, \mathbf{x_0}) \in E$. From the definition of a shadow graph, $h(\mathbf{x}_i) - h(\mathbf{x}_{i+1}) \geq W((\mathbf{x_i}, \mathbf{x_{i+1}}))(i = 0, \ldots, n-1)$ and $h(\mathbf{x}_n) - h(\mathbf{x}_0) \geq W((\mathbf{x_n}, \mathbf{x_0}))$. Therefore,

$$\sum_{i=0}^{n-1} W((\mathbf{x_i}, \mathbf{x_{i+1}})) + W((\mathbf{x_n}, \mathbf{x_0})) \leq \sum_{i=0}^{n-1} (h(\mathbf{x}_i) - h(\mathbf{x}_{i+1}))$$
$$+ (h(\mathbf{x}_n) - h(\mathbf{x}_0)) = 0. \qquad \square$$

The transitive closure of a shadow graph G with cycles is still well-defined because negative cycles do not interfere with the objective to seek paths with maximum accumulated weights according to the definition. The resulting graph G^c can still be used for obtaining an upper bound of the height for any point in $V - V_t$. Since there may be negative edges pointing from nodes in $V - V_t$ to some nodes in V_t, these edges can be used for obtaining a lower bound for some nodes in $V - V_t$. Since it is not guaranteed that there is an edge from

each node in $V - V_t$ to some node in V_t given a sparse set of images, we can only obtain lower bounds for a subset of nodes in $V - V_t$. And these lower bounds may appear useful in combination with other surface reconstruction techniques.

2.1. Shadow Graph Construction from Real Images

We typically use a single intensity threshold to perform shadow detection in real images. The shadow intensity threshold can either be chosen by the user or automatically determined. Shadow segmentation schemes will be discussed in Section 3.3.

When dealing with real images with noise, shadow thresholding cannot be expected to be error free. Inaccurate shadow segmentations may result in cycles in the induced shadow graphs. Since cycles can lead to the contradiction in the proof of Proposition 1, we must convert a cyclic graph into an acyclic one by removing some of the edges in the graph. Since we would like to make the least amount of change to the graph, a sensible criterion for an optimal conversion is that the total accumulated weight for the removed edges should be minimized. This is the *maximum acyclic subgraph problem*: given a directed graph $G = (V, A), V = 1, \ldots, n$, with arc weights $w_{ij} > 0, (i, j) \in A$, find a subset $A' \subset A$ such that $G' = (V, A')$ is acyclic and $w(A') = \sum_{(i,j) \in A'} w_{ij}$ is maximized. This problem is NP-hard (Karp, 1972).

To obtain an efficient approximation solution for this problem, we adopt the permutation-based algorithms in Hassin and Rubinstein (1994). A permutation π of $1, \ldots, n$ induces an acyclic subgraph $G_\pi = (V, A_\pi)$, where $A_\pi = \{(i, j) \in A \mid \pi(i) < \pi(j)\}$. Every maximal acyclic subgraph of G is induced by some permutation because one can always renumber the vertices of an acyclic graph so that each arc (i, j) in it satisfies $i < j$, and if the graph is maximal then it is the one induced by this permutation. A simple prototype algorithm (Hassin and Rubinstein, 1994) that generates a permutation, π, such that $w(A_\pi) \geq 0.5w(A)$, is given below. For $i \in V, S \subset V$, let $w_i^{\text{in}}(S) = \sum_{j \in S} w_{ji}, w_i^{\text{out}}(S) = \sum_{j \in S} w_{ij}$.

1. Set $S = V, l = 1, u = n$.
2. Choose $i \in S$. Set $S \leftarrow S - \{i\}$. If $w_i^{\text{in}}(S) \leq w_i^{\text{out}}(S)$, set $\pi(i) = l, l \leftarrow l + 1$. If $w_i^{\text{in}}(S) > w_i^{\text{out}}(S)$, set $\pi(i) = u, u \leftarrow u - 1$.
3. If $u \geq l$ go to Step 2. Else, stop and output π.

The various approximation algorithms in Hassin and Rubinstein (1994) can always produce an acyclic subgraph, but tend to remove more edges than necessary. After applying a permutation-based algorithm, for each of the removed edges, we still run a depth-first search to check whether the graph is still acyclic once the edge is inserted back into the graph. These two steps together lead to a polynomial time approximation that removes the least number of edges.

2.2.   Comparison with Shadowgrams

Shadow information can also be described using another representation known as a *shadowgram* (Kender and Smith, 1987; Daum and Dudek, 1998). Let us first look at shadowgrams for a one-dimensional terrain-like surface illuminated by a collimated light source parameterized by a single angle θ. As shown in Fig. 2(a), the shadowgram is a binary function defined on the angle, θ, and the spatial dimension of the terrain, x. A white entry in the shadowgram indicates that image pixel x would be lit when the light source were at angle θ, while a black entry indicates that it would be shadowed. It was shown in Kender and Smith (1987) that the shadowgram generated by a terrain-like surface can be completely described by two curves bounding the black regions: θ^+ and θ^-. For a specific point x, $\theta^+(x)$ and $\theta^-(x)$ represent the lighting angles at which x is lit for the first and last time, respectively. Therefore, $\theta^+(x) < \theta^-(x)$. If the light source travels from horizon to horizon, it is possible that one of these angles

might be equal to 0 or π. $\theta^-(x) - \theta^+(x) < \pi$ unless the point in question is a global maximum of the scene. Because one has this guarantee, it is possible to reconstruct the surface by integrating θ^+ and θ^-.

When the one-dimensional surface is non-terrain-like, the shadowgram possesses not only two curves, but also some white holes where one would expect darkness if the surface were a terrain (Fig. 2(b)). Here $\theta^+(x)$ and $\theta^-(x)$ are not defined as first and last lighting curves, but rather as the envelope of the shadowgram which lies closest to the middle horizontal line defined by $\theta = \frac{\pi}{2}$. It is shown in Langer et al. (1995) that using these two curves to reconstruct the surface will produce the surface's generated terrain which is defined to be its upper envelope. Furthermore, the holes in the shadowgram may be used to carve pieces out of the generated terrain for reconstructing some of the hidden parts of the surface.

In a general situation, a surface in a 3D space is parameterized by two variables, and the lighting direction is defined by two angles. The shadowgram thus becomes a four-dimensional function defined on all these four variables. Although shadowgrams can represent non-terrain-like surfaces, they are much more complicated than shadow graphs because they explicitly keep lighting directions in the representation. If we count the number of edges in a shadow graph, there is at most one incident edge at each pixel of an image because a pixel can be shadowed at most once by another pixel in the same image. Thus, the number of edges in a shadow graph is $O(kn^2)$ where k is the number of images which the shadow graph is constructed from and n^2 is the size
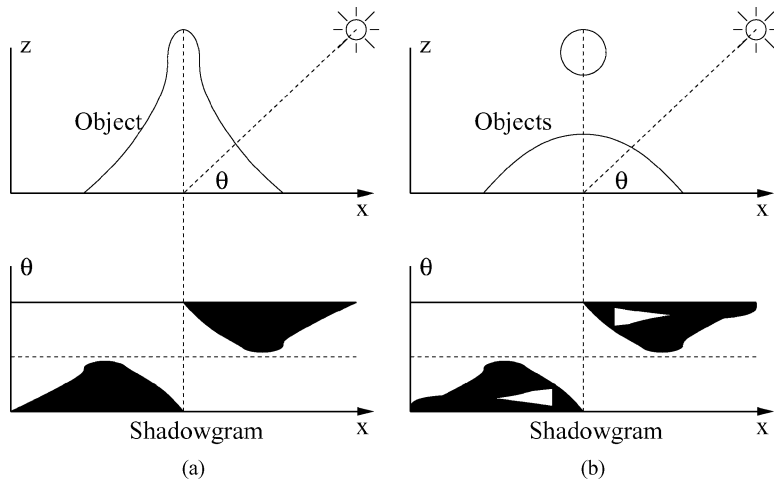

Figure 2.   (a) A shadowgram for a one-dimensional terrain-like surface; (b) a shadowgram for a one-dimensional non-terrain-like surface.

of each image. Recovering a height field from a shadow graph thus becomes tractable.

2.3. Experiments

We chose three synthetic examples to verify the capability of shadow graphs for surface reconstruction. In these examples, the height fields are reconstructed from shadows only. We also generated synthetic images from the height fields. The first dataset is a recovered height field of a real plaster sample using the approach presented in Liu et al. (2001). This height field serves as the ground truth to test the algorithm in this paper although we do not know the accuracy of this dataset. Sample input images are shown in Fig. 3(a) and synthetic images from the recovered height field are shown in Fig. 3(b). The second dataset is an artificial scene with four pyramids shown in Fig. 4(a). The pyramids have different height and orientations. Synthetic images from the recovered height field are shown in Fig. 4(b). The third dataset is a face model shown in Fig. 5(a). All the height fields used for synthetic rendering were recovered from 48 input images through the use of shadow graphs. The first three synthetic images in each example have the same illumination directions as the three input images shown above them. The last synthetic image has a novel illumination direction.

Table 1. The mean error of the height field, recovered from a shadow graph only, decreases significantly when an increasing number of images with shadows are used to compute the shadow graph. The first row shows the number of images used. The remaining rows show the decreasing errors for three examples. The errors are measured in pixels.

	4	8	12	16	24	48
Face	23.291	22.695	20.598	11.474	11.136	5.729
Plaster	8.261	6.510	6.168	6.062	5.970	4.480
Pyramids	26.789	22.621	19.090	18.194	14.873	7.950

Since the original height fields are known, we also verified that the accuracy of the recovered height fields increases with an increasing number of input images, as shown in Table 1. Since this is an asymptotic behavior predicted by Proposition 2, in general, a very large number of images are necessary to obtain an accurate reconstruction.

In consideration of this asymptotic behavior, the results shown in Figs. 3–5 are fairly reasonable. The overall shapes of the objects have been recovered and the relative height among different objects are also recovered according to the synthetic renderings and the cast shadows they have although the background plane in the pyramid and face models tends to have larger errors than the foreground objects due to the flatness of the background and a lack of
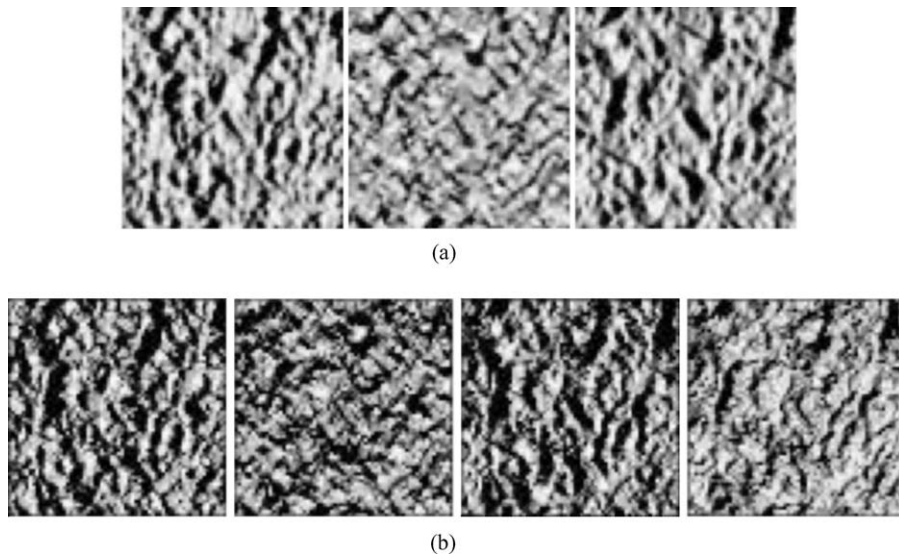

(a)

(b)

Figure 3. (a) Sample input images for a plaster material sample. (b) Synthetic images of a height field recovered from a shadow graph only. 48 images were used to construct the shadow graph. The rightmost image is a synthetic rendering of the recovered height field illuminated from a novel lighting direction.
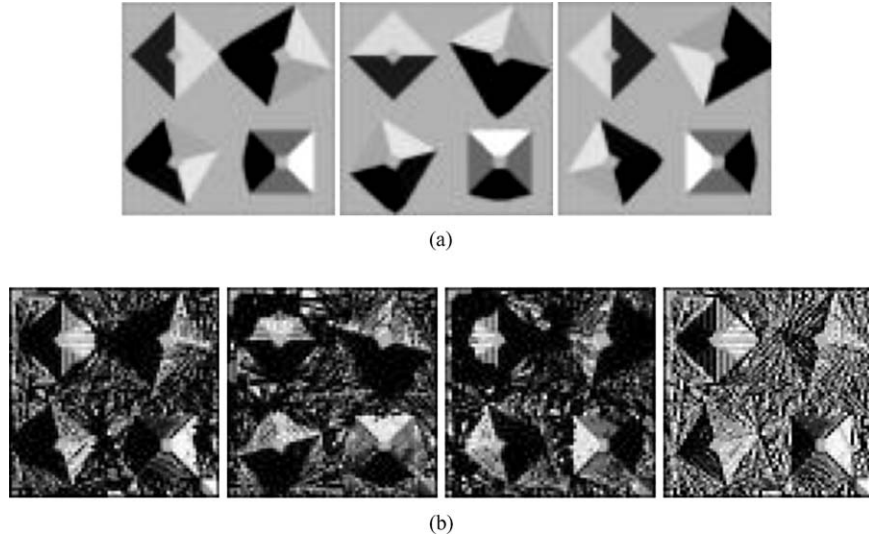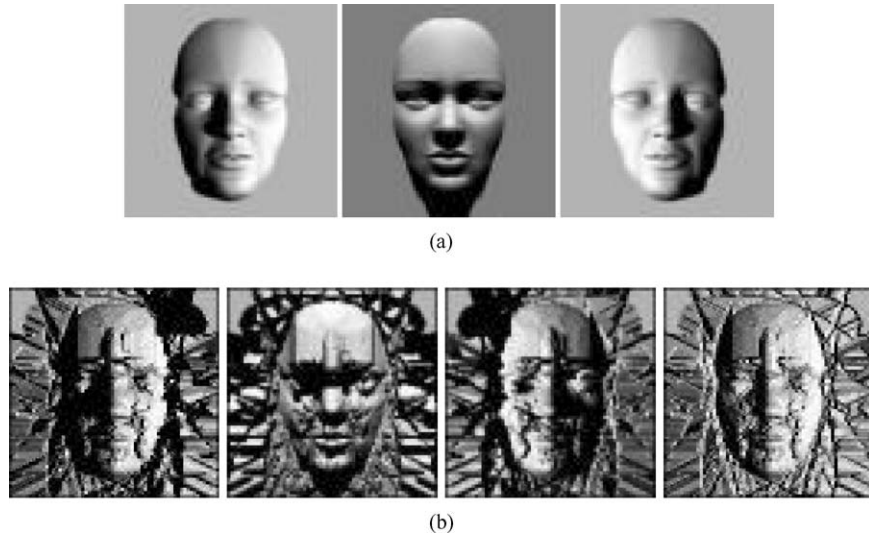
(a)

(b)

Figure 4. (a) Sample input images for a pyramid scene. (b) Synthetic images of a height field recovered from a shadow graph only. 48 images were used to construct the shadow graph. The rightmost image is a synthetic rendering of the recovered height field illuminated from a novel lighting direction.

(a)

(b)

Figure 5. (a) Sample input images for a face model. (b) Synthetic images of a height field recovered from a shadow graph only. 48 images were used to construct the shadow graph. The rightmost image is a synthetic rendering of the recovered height field illuminated from a novel lighting direction.

self-shadowing. However, the most noticeable artifacts are the numerous "chisel marks". This is because we only adopt shadow constraints without any form of regularization to smoothe the surfaces. The shadow constraints are inequalities most of the time, and only become equalities at the shadow boundaries. Therefore, the rough height field is just an upper enve-

lope of an underlying smooth surface. To some extent, shape-from-shadow recovers a surface by carving. Because of the "chisel marks", the height fields are not visually appealing. In the next section, we are going to explore a hybrid approach that can reconstruct height fields that are both visually and numerically accurate.

3. Height Field Reconstruction from Integrated Shadow and Shading Constraints

Recovering a height field from shadow graphs alone needs a large number of images. How can we reduce the number of input images? Given a sparse set of images with known lighting directions, we would like to recover the underlying height field using both shadow and shading constraints. As we have seen, shadows impose explicit constraints over surface height values, but shadows from sparse images are usually not sufficient if applied alone. On the other hand, shading information imposes additional constraints over normal orientations. Thus, it would be natural to integrate shadow constraints with shading information during surface reconstruction. We are going to explore two options where shadow constraints are considered either as soft constraints or as hard constraints. We first present an algorithm for soft shadow constraints. It is then generalized to enforce hard shadow constraints.

3.1.   Enforcing Shadow Constraints with Penalty Terms

Traditional shape-from-shading recovers surface normal orientations first followed by a numerical integration step to obtain a height field (Terzopoulos, 1988). This two-step approach is incompatible with shadow constraints which are directly expressed in terms of height values. Since shape-from-shading is not the focus of this paper, we adopt the direct height from shading algorithm in Leclerc and Bobick (1991) as the base for solving shading constraints. Since this technique computes a height field directly rather than through surface normals, it is relatively easy to incorporate shadow constraints and enforce surface upper/lower bounds from the previous section. The shape-from-shading problem is formulated to minimize the following cost function in Leclerc and Bobick (1991):

$$E_1 = \sum_{i,j} \left[\alpha(\rho R(p_{ij}, q_{ij}) - I(i, j))^2 + \lambda \left(u_{ij}^2 + v_{ij}^2 \right) \right] \tag{6}$$

where ρ is the surface albedo, I is the observed image intensity, p_{ij} and q_{ij} are not variables as in traditional shape-from-shading, but symmetric first-order finite differences of the surface height field $\{h_{ij}\}$ which is the unknown in this height-from-shading formulation, u_{ij} and v_{ij} are symmetric second-order finite dif-

ferences of the surface height field, α and λ are two constant coefficients, and R is the Lambertian reflectance model. The first term in Eq. (6) corresponds to the photometric error term. And the second is a regularization term on the smoothness of the surface.

The shape-from-shading problem is underconstrained and the regularization term is crucial to obtain a reasonable solution. However, 3D textures typically have high-frequency small-scale details. Being overly dependent on the regularization term would hamper the accurate recovery of these details. Therefore, we decide to use multiple images to overconstrain the problem as in photometric stereo (Woodham, 1989). Obviously, shadowed pixels need to be ignored in shape-from-shading. The above height-from-shading formulation can be easily generalized to accommodate multiple input images and shadow masks as follows.

$$E_2 = \alpha \sum_{k} \sum_{i,j} l_{ij}^k (\rho_{ij} R^k(p_{ij}, q_{ij}) - I^k(i, j))^2 \\ + \lambda \sum_{i,j} \left(u_{ij}^2 + v_{ij}^2 \right) \tag{7}$$

where $I^k(i, j)$ represents the k-th input image with corresponding reflectance map R^k, l_{ij}^k is a binary shadow mask indicating whether pixel (i, j) in the k-th image is lit by the light source or not, and ρ_{ij} is the pixelwise surface albedo. Although this treatment is similar to photometric stereo, it solves the height field directly instead. With multiple images, the regularization term becomes much less important, and its weight can be set close to zero. However, it is still helpful at underconstrained pixels that are lit in less than three different images. In principle, when the surface is static and Lambertian, both the height field $\{h_{ij}\}$ and albedo map $\{\rho_{ij}\}$ can be solved simultaneously from three or more input images with non-coplanar illumination directions. In practice, we found that it is more accurate to solve for the albedo map first using photometric stereo, and recover the height field using Eq. (7) with a known albedo map. Note that the normals recovered from photometric stereo are discarded since Eq. (7) can directly recover the height field. Because only the Lambertian reflectance model is explicitly considered, specular highlights are treated as outliers since they are quite concentrated and usually cover a small number of pixels. With multiple input images, each pixel has multiple intensity measurements and the majority of the intensities observe Lambertian model well. The rest of the intensities are labeled as outliers and iteratively reweighted (Hampel et al., 1986) during minimization.

To further incorporate the constraints in Eqs. (1) and (4) into the above formulation, we notice that the constraints have the same form which looks like

$$h_{ij} - h_{i'j'} \geq d_{iji'j'}. \qquad (8)$$

To enforce these kind of inequalities in a gradient-based minimization method, a differentiable half-sided parabola is adopted as a penalty function,

$$S(i, j, i', j') = [\min(0, h_{ij} - h_{i'j'} - d_{iji'j'})]^2. \qquad (9)$$

The penalty functions for all the inequalities and equalities can be inserted as additional terms into Eq. (7). The new cost function for surface reconstruction is given as follows.

$$\begin{aligned} E_3 = &\alpha \sum_k \sum_{i,j} l_{ij}^k (\rho_{ij} R^k(p_{ij}, q_{ij}) - I^k(i, j))^2 \\ &+ \lambda \sum_{i,j} \left(u_{ij}^2 + v_{ij}^2\right) \\ &+ \beta \sum_k \left(\sum_{m_k} S^k\left(i_{m_k}, j_{m_k}, i'_{m_k}, j'_{m_k}\right) + T_k \right) \quad (10) \end{aligned}$$

where m_k is the index of the inequality constraints from the k-th image, $S^k(i_{m_k}, j_{m_k}, i'_{m_k}, j'_{m_k})$ represents the actual penalty terms contributed by the k-th image, and T_k represents the collection of penalty terms for the equality constraints associated with shadows, such as those in Eqs. (2) and (3). α, β and λ are weights used to indicate the relative importance of each term. They are currently set up by the user. In our experiments, we use iterative minimization algorithms and require $\alpha = 1 - \lambda$. λ is initialized to 0.1 and divided by a constant factor after each iteration to phase out the regularization term. We set $\beta \geq 1$ to emphasize the importance of shadow constraints.

All the above three cost functions are differentiable and can be minimized by the standard conjugate gradient algorithm (Press et al., 1988). Since the cost functions have local minima, it is important to start from a good initialization to converge to the correct solution. In practice, multiresolution minimization can perform more efficiently and accurately. When the input images have a high resolution, an image pyramid can be built for each of them and the solution obtained from a lower resolution can serve as a good initial solution for a higher resolution.

3.2. Enforcing Upper and Lower Bounds

In the formulation in Section 3.1, shadow constraints are enforced as soft constraints by using penalty terms

in an original height-from-shading algorithm. It is not guaranteed that all constraints are satisfied. Sometimes, it is more desirable to consider shadow constraints as hard constraints since they are less sensitive to specular reflection and albedo variations, and to consider shading constraints as soft ones since a small amount of deviation in shading is not very noticeable. Directly using the shadow and antishadow constraints in a constrained optimization framework is computationally expensive since the number of such constraints can be much larger than the number of pixels. The upper and lower bounds derived from the shadow graph in Section 2 can better serve as constraints during optimization since the number of such bounds is only $O(n^2)$ which is on the same order of magnitude as the number of pixels.

The upper and lower bounds can be estimated as follows. (Note that the height of the nodes in the set V_t is unknown at the beginning, and they can be estimated from a solution of the height field from Section 3.1.)

1. Obtain an initial estimation of the height value for each point by minimizing Eq. (10);
2. Adjust the initial height values of the nodes in V_t to satisfy all the antishadow constraints among them as in the following convergent procedure;

    (a) fix the height of the highest point in V_t;
    (b) loop through the rest of the points and check whether the considered point is in the shadow of some other point in V_t because of the violation of a antishadow constraint; if so, raise the considered point to the minimum height that can eliminate the violation.

3. Calculate the upper and lower bounds for nodes in $V - V_t$ from the transitive closure G^c.

To enforce the upper and lower bounds, our complete algorithm still takes an initial solution of the height field from minimizing Eq. (10). However, there are multiple possibilities to improve this initial solution:

1. For each point, if it is higher than its upper bound, push it down to the upper bound; if it is lower than its lower bound, raise it to the lower bound.
2. Use a constrained optimization algorithm to enforce the upper and lower bounds.
3. Fix a subset of the adjusted points from the first step and minimize Eq. (10) with those fixed points as additional boundary conditions; alternate adjustment and minimization (with a few additional fixed points every iteration) until all the bounds are satisfied.

The first scheme chooses to satisfy all the hard constraints by using brute force and ignoring all the shading constraints, therefore tends to have unnatural discontinuities at those adjusted places. The second scheme chooses to apply a general-purpose constrained optimization algorithm to automatically and iteratively adjust the heights so that the bounds are satisfied at the end. Unfortunately, constrained optimization algorithms, such as sequential quadratic programming (SQP) and nonlinear bounded least-squares are usually computationally expensive on high-dimensional data such as images. Such algorithms repeatedly solve local quadratic programming problems which require the inversion of a matrix of the same size of a Hessian matrix. If the images have size $n \times n$, the number of unknowns in the optimization is n^2, and the size of a Hessian is $n^2 \times n^2$.

The last scheme chooses to adapt unconstrained optimization algorithms so that they allow a part of the variables to be fixed. To achieve that, we can simply set the corresponding derivatives to be zero. We fix a few additional points within their bounds before unconstrained minimization takes place in every iteration. This can satisfy all the bounds in a finite number of iterations since we only try to recover height values at a finite number of points (pixels). An intuitive illustration is given in Fig. 6.

In practice, we chose the last scheme with some additional details. After initialization, the height values of the nodes in V_t, and the upper and lower bounds are fixed in all iterations. In every iteration, we subtract the upper bounds from the current estimation of the height field to obtain a difference field. Then the set of local maxima in the difference field are located. Those points corresponding to the local maxima are lowered to their corresponding upper bounds and fixed thereafter. Other points higher than their upper bounds are also modified to be lower than their upper bounds, but not fixed. The same procedure is repeated for lower bounds before the unconstrained minimization in Eq. (10) takes

place once again with the newly fixed points as additional boundary conditions. We hope that the shading constraints solved during minimization can automatically adjust the neighborhoods of those fixed points so that there will be much less violated bounds in the next iteration. This can also avoid having many unnatural discontinuities since the minimization procedure serves as a smoothing operator by considering all constraints simultaneously.

Since we adopt unconstrained optimization during each iteration by simply using zero derivatives for those fixed variables, this custom-designed algorithm is much more efficient than those general-purpose constrained optimization algorithms. For example, the sequential quadratic programming software package (Fsqp software, http://gachinese.com/aemdesign/FSQPframe.htm) we have tested took two hours to finish one iteration on eight 64×64 images on a Pentium III 800 MHz workstation while our algorithm only took less than one minute each iteration. The actual number of iterations to achieve convergence varies with examples. Usually dozens of iterations are needed before convergence.

3.3. Shadow Segmentation

Since we exploit shadow constraints during surface reconstruction, it is important to obtain accurate shadow segmentation. There are two different types of shadow boundaries. According to Fig. 1(a), the shadow boundary at the occluder x_0 is called the self shadow boundary while the shadow boundary at x_1 is called the cast shadow boundary. It is straightforward to see that a cast shadow boundary generates an intensity discontinuity since we use collimated light sources. However, the intensity across a self shadow boundary over a differentiable surface is continuous since the surface normal varies smoothly and gradually points away from the light source. The shadow boundary is at the location
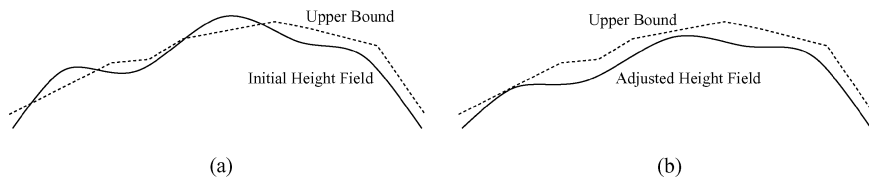

(a)   (b)

Figure 6. (a) Some parts of the height field recovered from minimization may exceed the upper bound; (b) We need to globally adjust the initial height field to maintain its original smoothness instead of simply clipping it against the upper bound.

where the surface normal becomes perpendicular to the lighting direction. The continuous change of intensity makes a self shadow boundary hard to detect. For 3D textures, this problem is less severe since 3D textures have high-frequency geometric details which make self shadow boundaries easier to localize.

There are two additional factors that affect the sharpness of a shadow boundary. The first one is the sensor's ground noise level. It is impossible to obtain perfectly black shadow regions since every camera has ground noise and the signal to noise ratio drops dramatically

when the incident intensity becomes close to zero. The second factor is mutual interreflections in the shadow regions. Shadow regions receive a small amount of indirect illumination from the surrounding environment to make themselves slightly brighter. These two factors make it harder to accurately localize the shadow boundaries.

There have been significant amount of previous work on shadow detection (Branca et al., 2002; Prati et al., 2003) most of which focuses on detecting shadows from video sequences where either the object
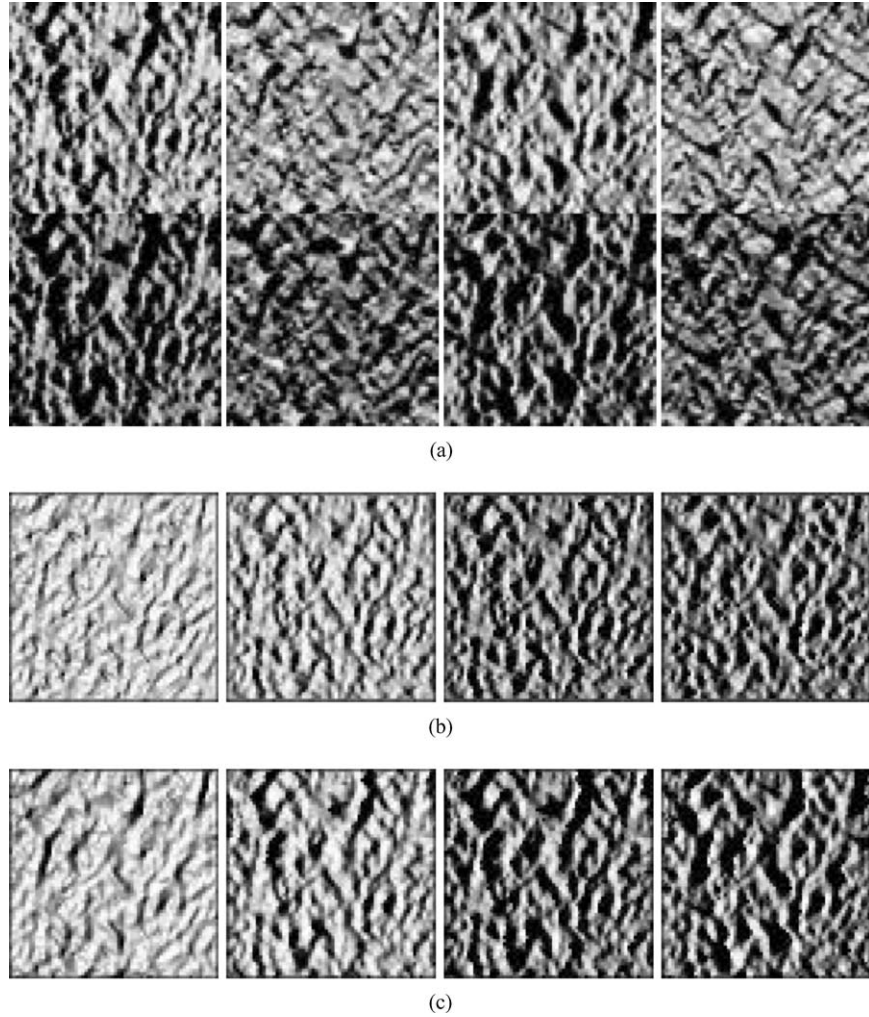

Figure 7. (a) Input images for the plaster material sample. The angle of elevation of the lighting directions in the top row is 45 degrees, the bottom row 30 degrees. (b) Synthetic images of a height field recovered from shape-from-shading only. (c) Synthetic images of the height field recovered from our hybrid method. In terms of illumination directions, the first image has a novel direction and each of the other three in (c) has a corresponding image in (a). The second image corresponds to the first one of the first row in (a); the third image corresponds to the first one of the second row; and the last image corresponds to the third one of the second row. Every image in (c) also has the same illumination direction as their corresponding image in (b).

casting the shadow is moving or the camera is moving. Temporal differencing and background subtraction have been commonly exploited in the literature. Prati et al. (2003) provides a good evaluation and comparison among various algorithms. In terms of shadow detection in still images, as is the case in this paper, intensity thresholding is the dominant approach. The specific threshold for the shadow pixels is typically user-defined. More advanced image segmentation techniques, such as region growing and graph partitioning (Shi and Malik, 2000), tend to exploit spatial coherence to group nearby pixels together. Since we try to recover the height fields of 3D textures with high spatial frequencies, maintaining spatial coherence is not desirable. In addition, since the images we use for material samples are taken in a well controlled environment without ambient illumination, the intensities of shadowed and nonshadowed pixels are reasonably separated. Therefore, thresholding is both effective and convenient.

To achieve more accurate pixel classification than simple thresholding, we designed two schemes to
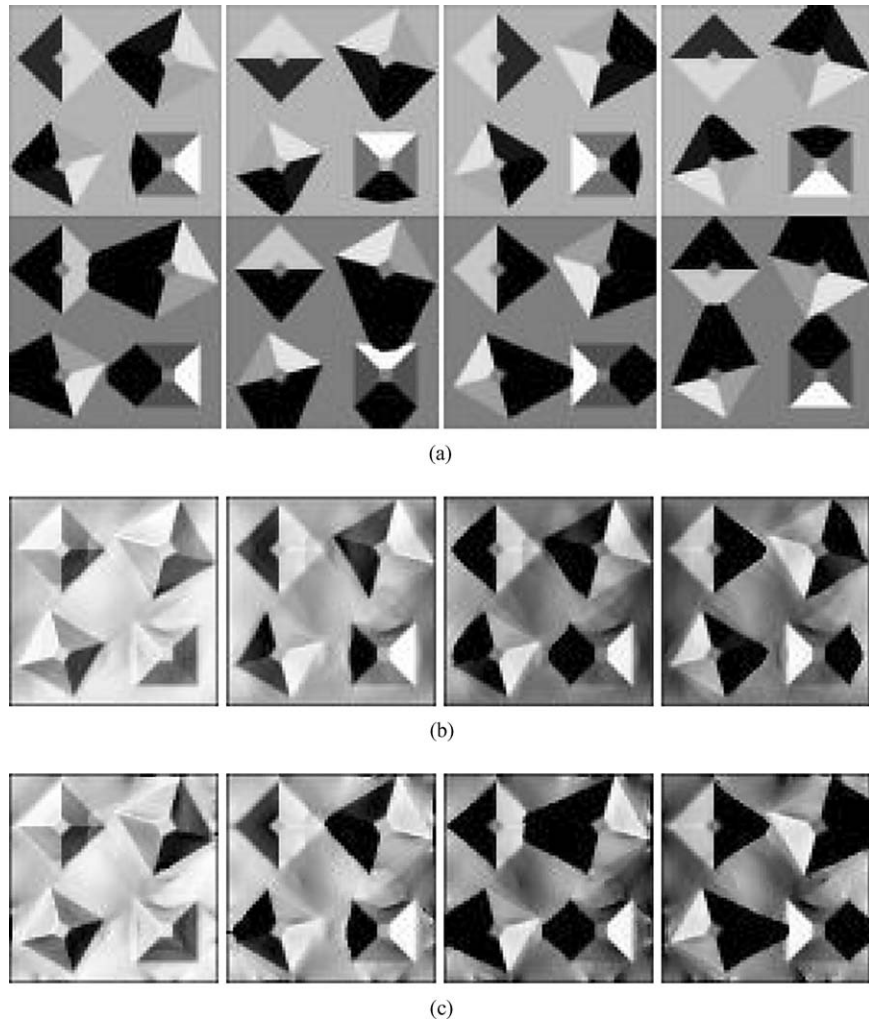

Figure 8. (a) Input images for the pyramid scene. The angle of elevation of the lighting directions in the top row is 45 degrees, the bottom row 30 degrees. (b) Synthetic images of a height field recovered from shape-from-shading only. (c) Synthetic images of the height field recovered from our hybrid method. In terms of illumination directions, the first image has a novel direction and each of the other three in (c) has a corresponding image in (a). The actual correspondence is the same one used for Fig. 7. Every image in (c) also has the same illumination direction as their corresponding image in (b).

perform thresholding more carefully. The first scheme needs the user to specify two intensity thresholds: a lower threshold T_1 and a higher threshold T_2. They represent conservative thresholds for shadowed regions and illuminated regions, respectively. During surface reconstruction, all the pixels with intensity below T_1 are regarded as shadowed pixels and shadow constraints are imposed while pixels with intensity above T_2 are regarded as illuminated pixels and shading constraints are imposed. In this scheme, we simply do not make use of the pixels with intensity between the two thresh-

olds to avoid possible errors from inaccurate shadow segmentation. Of course, we assume that the ignored pixels in a subset of the images are actually considered in the rest of the images. Otherwise, we have to rely on the regularization term to make reasonable speculations on their height values.

In the second scheme, we consider the shadow threshold as an additional parameter that needs to be determined during surface reconstruction. It is still reasonable for the user to define a range $[T_1, T_2]$ for the shadow threshold. An automatic procedure is then
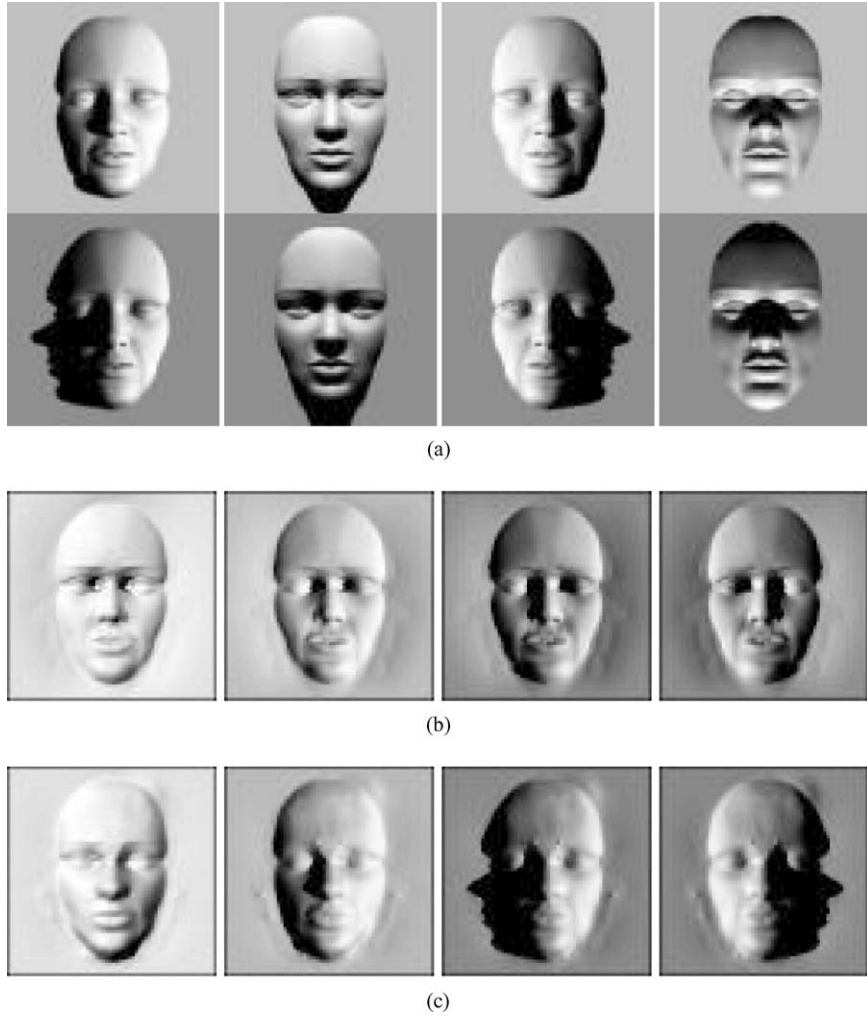

Figure 9. (a) Input images for the face model. The angle of elevation of the lighting directions in the top row is 45 degrees, the bottom row 30 degrees. (b) Synthetic images of the height field recovered from shape-from-shading only. (c) Synthetic images of the height field recovered from our hybrid method. In terms of illumination directions, the first image has a novel direction and each of the other three in (c) has a corresponding image in (a). The actual correspondence is the same one used for Fig. 7. Every image in (c) also has the same illumination direction as their corresponding image in (b).
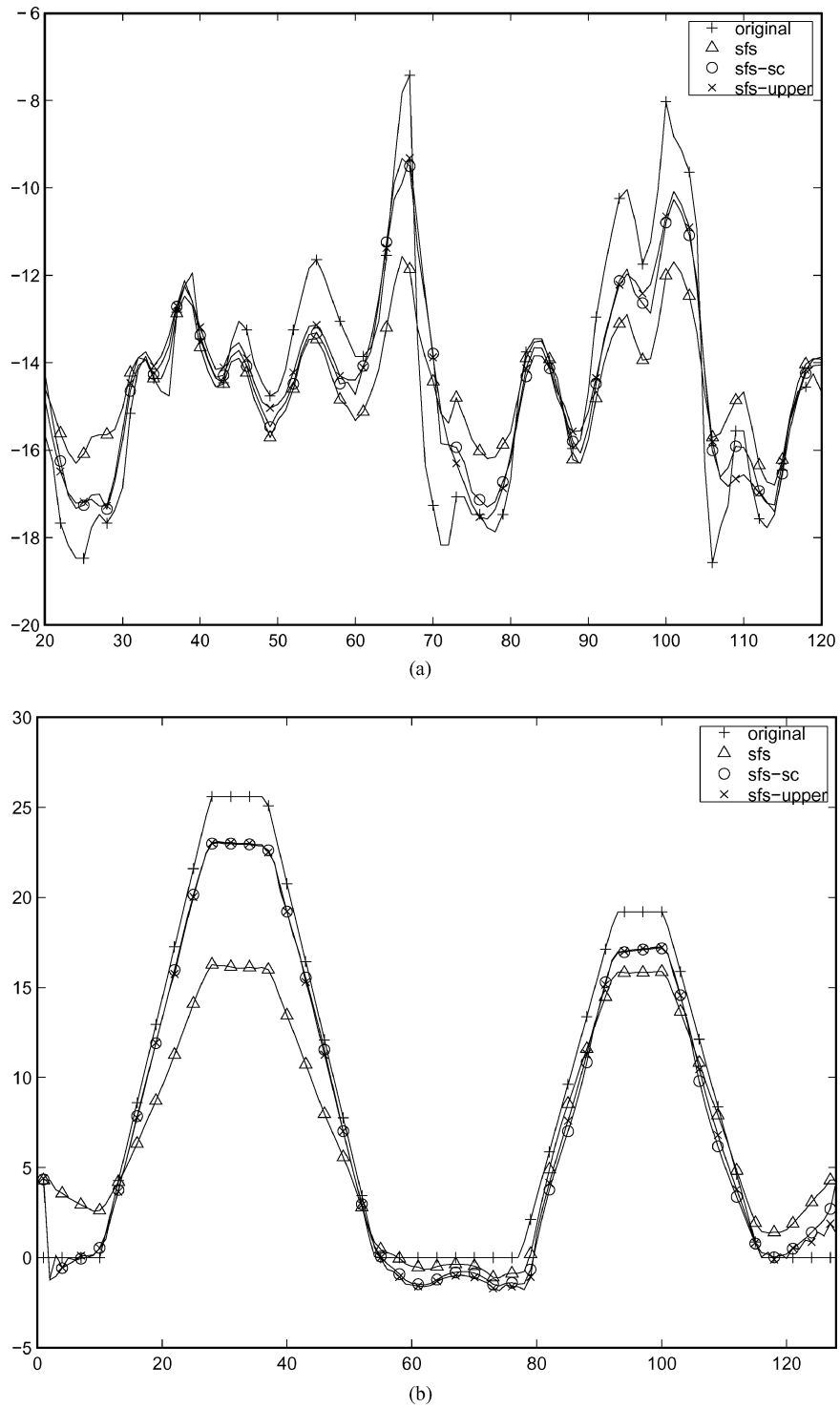
Figure 10. Comparison of the cross sections of four height fields: the ground truth is shown as 'original'; the one from minimizing Eq. (7) is shown as 'sfs'; the one from minimizing Eq. (10) is shown as 'sfs-sc'; and the one from enforcing bounds is shown as 'sfs-upper'. (a) Cross sections for the plaster sample. (b) Cross sections for the pyramid scene.

executed to search for the best threshold within the given range. The automatic procedure is based on the following heuristic: given the correct shadow threshold, most pixels will be correctly classified as either shadowed or illuminated pixels, correct constraints will be imposed on them during optimization, and the cost of Eq. (10) after minimization should be very low; when the shadow threshold is too low or too high, some pix-

els will be misclassified, incorrect constraints will be imposed on them, and the cost of Eq. (10) after minimization remains higher than it should be. According to the above heuristic, we iteratively adjust the shadow threshold using the one-dimensional Golden Section

Table 2. Comparison of three algorithms on three datasets using noise-free images: (i) shape-from-shading only by minimizing E_2 in Eq. (7), (ii) hybrid method by minimizing E_3 in Eq. (10), (iii) hybrid method by enforcing bounds as in Section 3.2.

	Plaster	Pyramids	Face
E_2	1.478, 1.934	2.481, 3.658	3.552, 4.416
E_3	1.099, 1.455	1.782, 2.242	1.508, 1.743
E_3+Bounds	1.005, 1.400	1.671, 2.198	1.367, 1.614

Each entry has a pair of error measurements for the recovered height field. The first one is the mean error while the second one is the RMS error. All numbers are given in the unit of a pixel.

Table 3. Comparison of three algorithms on three datasets using images with 5% noise: (i) shape-from-shading only by minimizing E_2 in Eq. (7), (ii) hybrid method by minimizing E_3 in Eq. (10), (iii) hybrid method by enforcing bounds as in Section 3.2.

	Plaster	Pyramids	Face
E_2	1.483, 1.940	2.544, 3.762	3.583, 4.452
E_3	1.055, 1.409	1.809, 2.268	1.500, 1.753
E_3+Bounds	1.039, 1.396	1.768, 2.210	1.493, 1.749

Each entry has a pair of error measurements for the recovered height field. The first one is the mean error while the second one is the RMS error. All numbers are given in the unit of a pixel.

Table 4. Comparison of accuracy and timing for the face model in terms of number of input images and number of levels in a multiresolution optimization.

Face	4	8	24	48
1 level	4.160, 36s	3.968, 80s	3.563, 277s	1.486, 569s
2 levels	3.529, 32s	3.268, 57s	2.938, 177s	1.334, 398s
3 levels	1.624, 26s	1.367, 48s	0.978, 173s	0.977, 383s

The first row shows the number of images. The remaining rows show the mean errors of the height fields and timing results simultaneously when image pyramids between one to three levels are used during optimization. The mean reconstruction errors are given in the unit of a pixel, while the timing results are given in seconds.
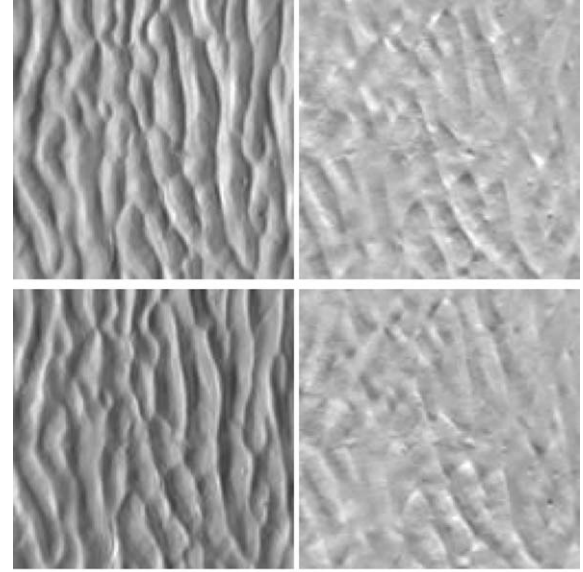
(a)
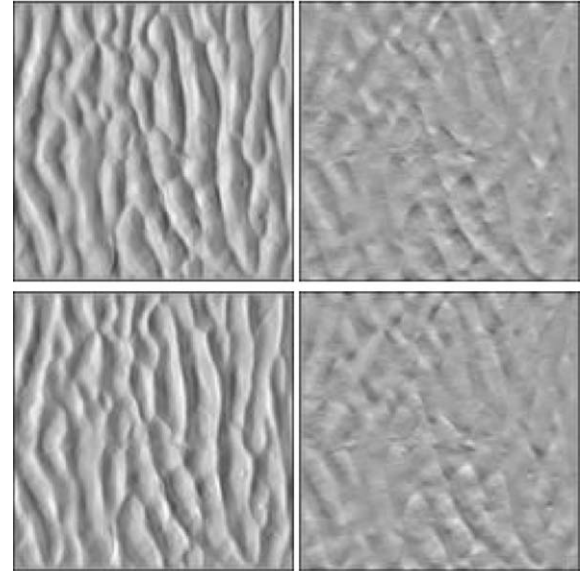
(b)

Figure 11. (a) Input images of a material sample from the Heriot-Watt PhoTex database. The angle of elevation of the illumination is 15 degrees. The azimuth angles are 0, 90, 180 and 270 degrees. (b) Synthetic images of the recovered height field illuminated from the same lighting directions as the input images.

search (Press et al., 1988). One iteration for shadow threshold adjustment involves the whole process of minimizing Eq. (10) with the current shadow threshold fixed. Thus, surface reconstruction and shadow segmentation can be carried out simultaneously.

3.4.   Experiments

We have tested our algorithms using integrated shadow and shading constraints on both synthetic and real imagery.

3.4.1. Synthetic Data.   We chose to use the same synthetic examples in Section 2.3 to verify the effectiveness of our algorithms. The first example is a 3D texture while the other two represent general height fields to indicate that our algorithms are not limited to 3D textures. Eight synthetic images were generated as input for each of the three datasets we chose. Four of them were lit from an angle of elevation of 45 degrees and the others were lit from 30 degrees to create images with significant amount of shadow. We apply both shape-from-shading and our hybrid algorithm to recover the
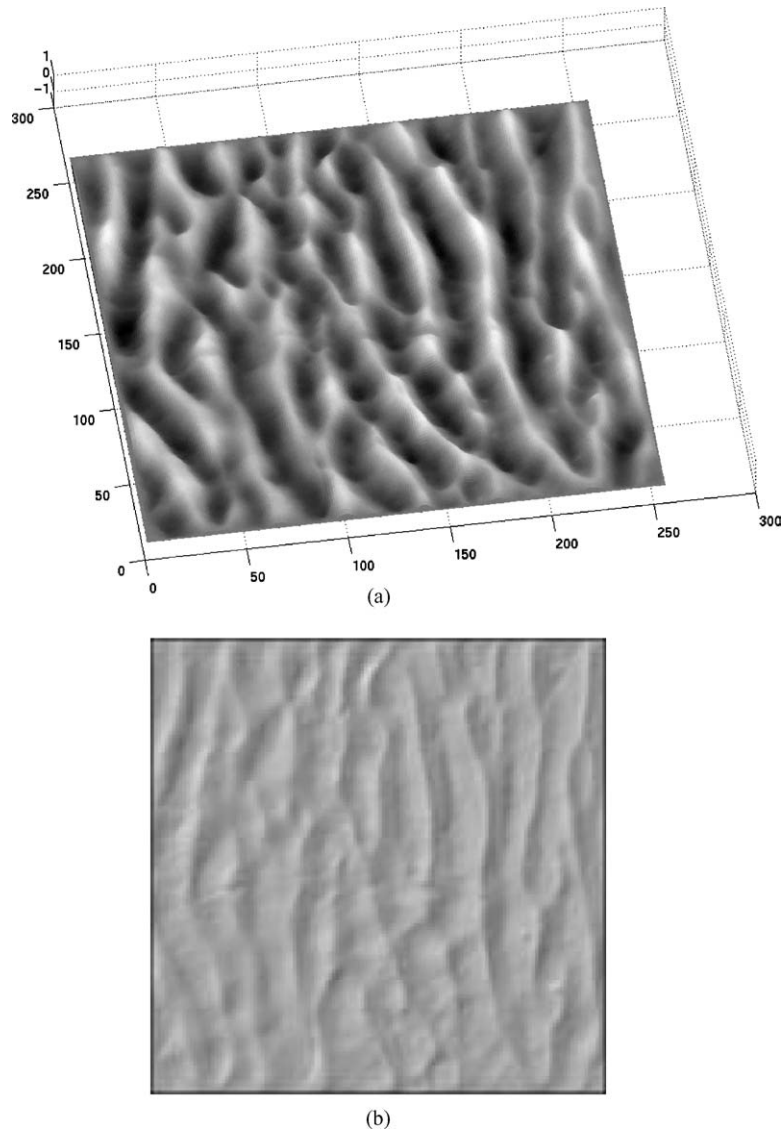

(a)

(b)

Figure 12.   (a) The recovered height field for the material sample in Fig. 11. The height field is visualized as a gray-scale image. (b) A synthetic image of the recovered height field illuminated from a novel lighting direction with an angle of elevation 45 degrees.

height field for each example. Since the original height fields are known, we compared the accuracy of the height fields recovered from both approaches based on error measurements. We also generated synthetic images for each example from the recovered height fields. Some of the synthetic images are lit from the same lighting directions as some of the input images to verify both shadowed and non-shadowed regions while the other synthetic images are lit from a novel lighting direction which is different from the ones for the input images to show that the recovered height fields can be useful for creating images with correct appearance from novel lighting conditions.

The input images for the plaster sample are shown in Fig. 7(a), and the synthetic images of the height fields recovered from shape-from-shading and our hybrid approach are shown in Fig. 7(b) and (c), respectively. The input images for the pyramid scene are shown in Fig. 8(a). The synthetic images of the recovered height fields are shown in Fig. 8(b) and (c). Figure 9 shows the input and synthetic images for the face model. In these examples, the solutions obtained from shape-from-shading do not preserve relative height differences among surface points very well. This is most noticeable from the synthetic images shown in Fig. 9(b) where the recovered face model looks too flat while the face model recovered by our hybrid approach is visually much more closer to the origial one. For the same reason, the synthetic images for the height fields recovered from shape-from-shading typically do not have sufficient amount of cast shadow. Our hybrid algorithm managed to enforce the shadow constraints and make the generated images look more similar to the input ones. Considering shape-from-shadow only, the results in Section 2.3 actually appear less flattened than the results from shape-from-shading which, however, tends to generate very smooth surfaces.

Figure 10 shows two comparisons of the cross sections. In each of the comparisons, there are four curves including the ground truth, the curve from minimizing Eq. (7), the curve from minimizing Eq. (10) and the curve from enforcing the height bounds. The results from minimizing Eq. (7) are not as good as the other two versions because it does not consider shadow constraints. In our examples, most points are lit from at least one lighting direction. The height field can be recovered from shape-from-shading or photometric stereo alone. However, the additional shadow constraints can definitely improve the accu-

racy of the results because shading only imposes local constraints, and shading-based techniques can introduce accumulated errors from pixel to pixel during integration while shadow constraints are very
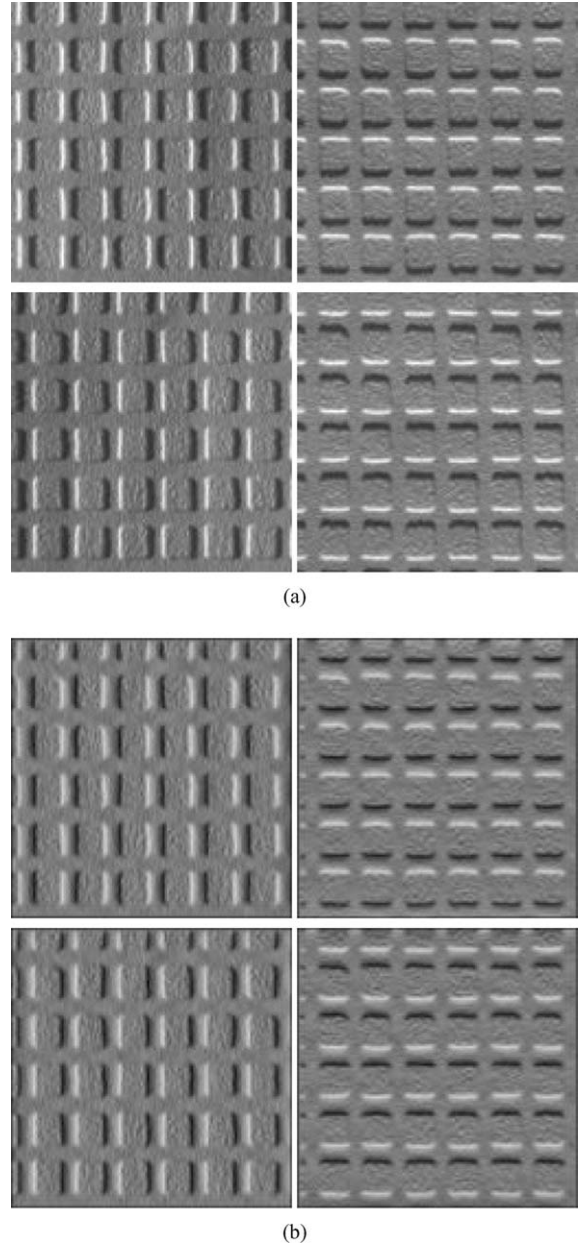

(a)

(b)

Figure 13. (a) Input images of a material sample from the Heriot-Watt PhoTex database. The angle of elevation of the illumination is 15 degrees. The azimuth angles are 0, 90, 180 and 270 degrees. (b) Synthetic images of the recovered height field illuminated from the same lighting directions as the input images.

good at directly enforcing long-range relative height constraints.

Tables 2 and 3 show the error measurements of the recovered height fields from three algorithms: the variant of shape-from-shading described in Eq. (7), the algorithm enforcing soft shadow constraints by minimizing Eq. (10), and the generalized algorithm enforcing hard height bounds. The latter two consistently performed better than shape-from-shading and the amount of improvement is significant. The numerical accuracy of the latter two algorithms does not differ significantly although the version enforcing hard bounds performed slightly better. In Table 3, a zero mean Gaussian noise with a standard deviation of 5% of the gray-scale range (0-255) is added to the original input images used for Table 2. The three algorithms all performed robustly
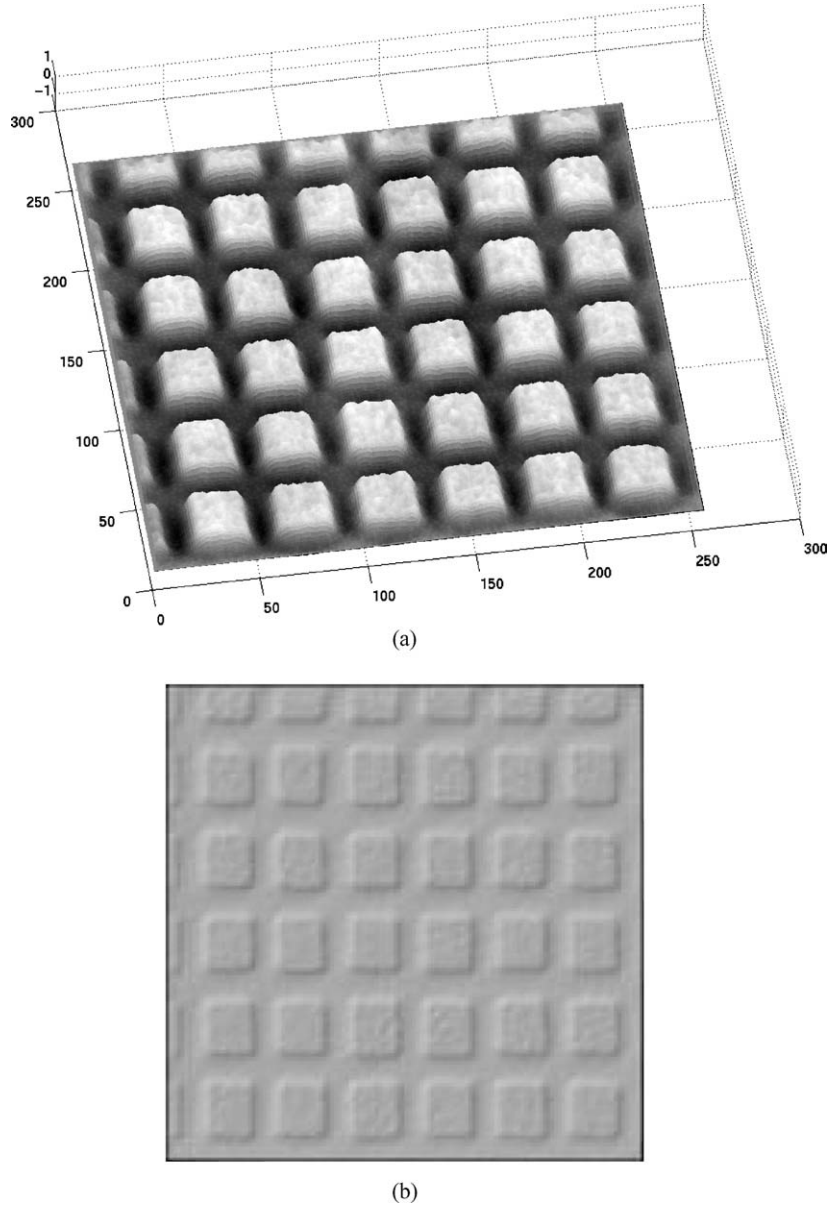

(a)

(b)

Figure 14.   (a) The recovered height field for the material sample in Fig. 13. The height field is visualized as a gray-scale image. (b) A synthetic image of the recovered height field illuminated from a novel lighting direction with an angle of elevation 45 degrees.

by only slightly increasing the errors in the recovered height fields. The amount of increased error is typically below 5%.

We implemented a multiresolution version of the hybrid algorithm and did a comparison on the face dataset using image pyramids between one to three levels. An image pyramid is built for each of the input images. The multiresolution optimization begins at the coarsest resolution and recovers a height field using the coarsest image from each image pyramid. The coarsest height field is then interpolated to initialize the height field at the next higher resolution. This step is repeated until the highest resolution has been reached. A convergence test is also performed simultaneously with the multires-

olution test. For a fixed number of levels, an increasing number of input images are used to recover the height field. The error measurement and timing result for each combination are shown in Table 4. In summary, the algorithm runs faster and produces more accurate results with an increasing number of resolutions. The algorithm runs slower but produces more accurate results with an increasing number of input images.

3.4.2. Real 3D Textures. We performed extensive experiments on real material samples captured in the Heriot-Watt PhoTex database (Dong and Chantler, 2002) and the Columbia-Utrecht *CUReT* database (Dana et al., 1999). We focus our experiments on those
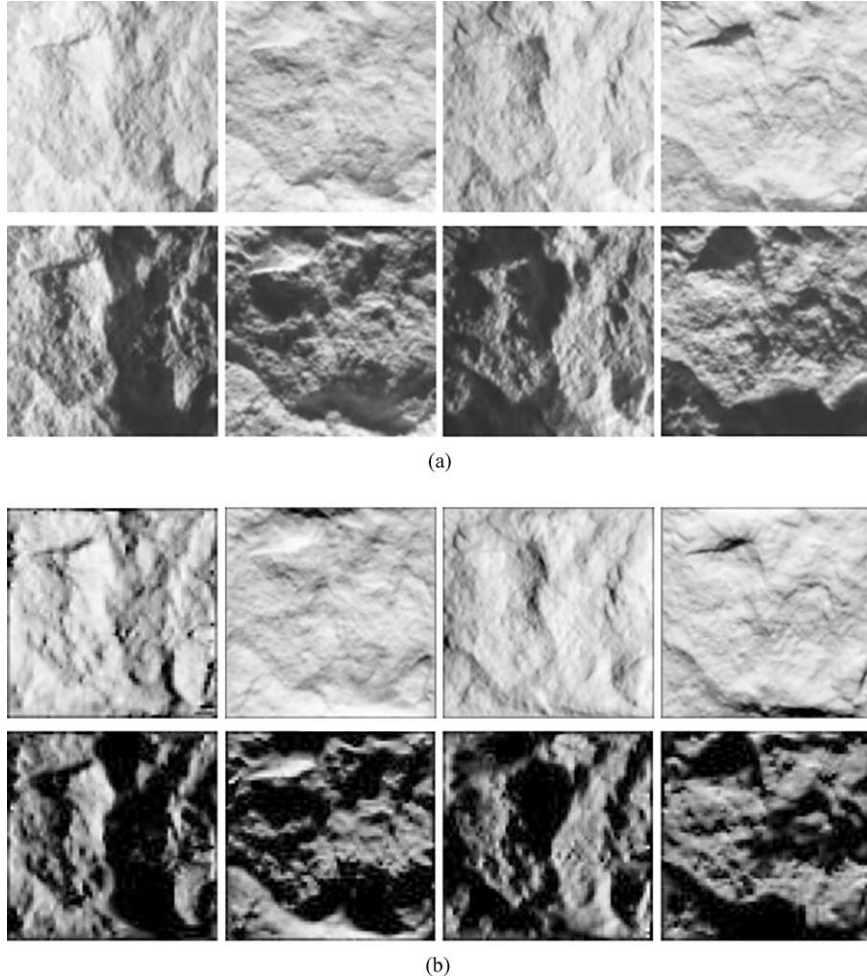

(a)

(b)

Figure 15. (a) Input images of a material sample from the Heriot-Watt PhoTex database. The illumination for the first row has an angle of elevation 45 degrees while the second row 15 degrees. (b) Synthetic images of the recovered height field illuminated from the same lighting directions as the input images.
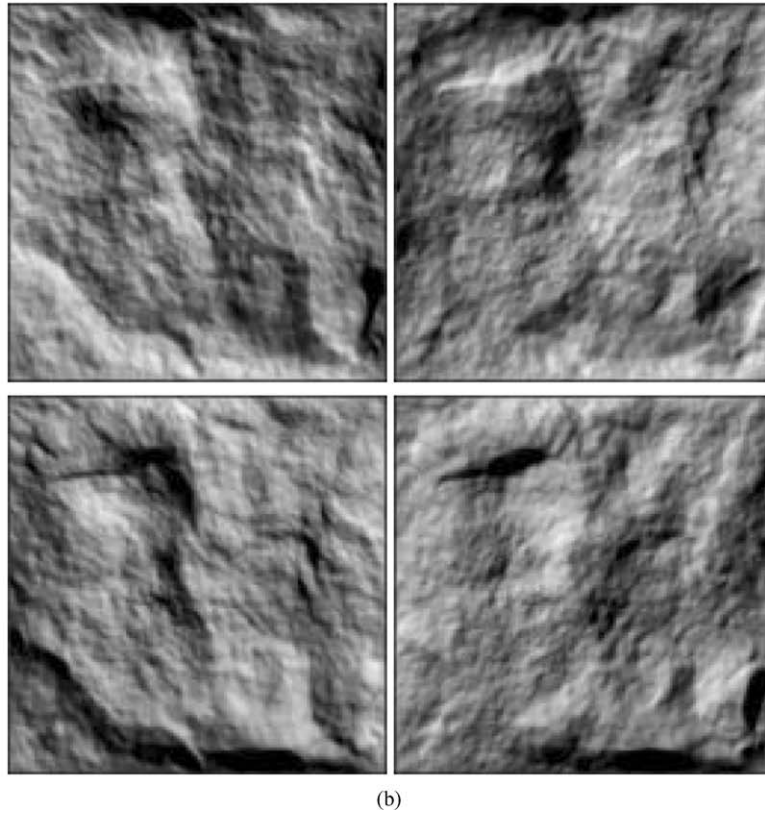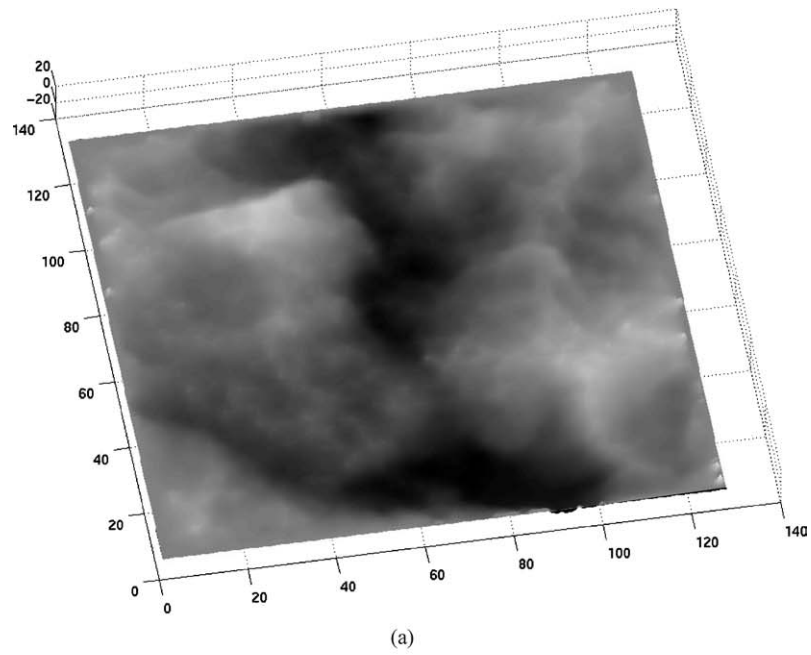
(a)

(b)

Figure 16.   (a) The recovered height field for the material sample in Fig. 15. The height field is visualized as a gray-scale image. (b) Synthetic images of the recovered height field illuminated from novel lighting directions with an angle of elevation 30 degrees. The azimuth angles are 45, 135, 225 and 315 degrees.

material samples with obvious height relief since explicit height field reconstruction may not be necessary for the rest of the material samples.

Figures 11–16 show some of the results on material samples captured in the PhoTex database. Only four or eight 256×256 input images were chosen for each material sample. Many of the input images have shadows. Because of the spatial height pattern on the material, images corresponding to different illumination directions may have dramatically different appearances (for example, Fig. 11). The height fields were recovered using our hybrid shape from shadow and shading approach. The shadow threshold for each sample was determined automatically within a user defined range using the second technique discussed in Section 3.3. For example, this threshold for the sample shown in Fig. 15 was found to be 92 out of 256 possible gray-scale levels. The reconstructed surfaces were then synthetically rendered using the original as well as novel illumination directions. From these synthetic images, we can see that interesting high-frequency details in the height fields have been preserved, and the recovered height fields can reproduce the dramatic appearance changes due to different lighting directions.
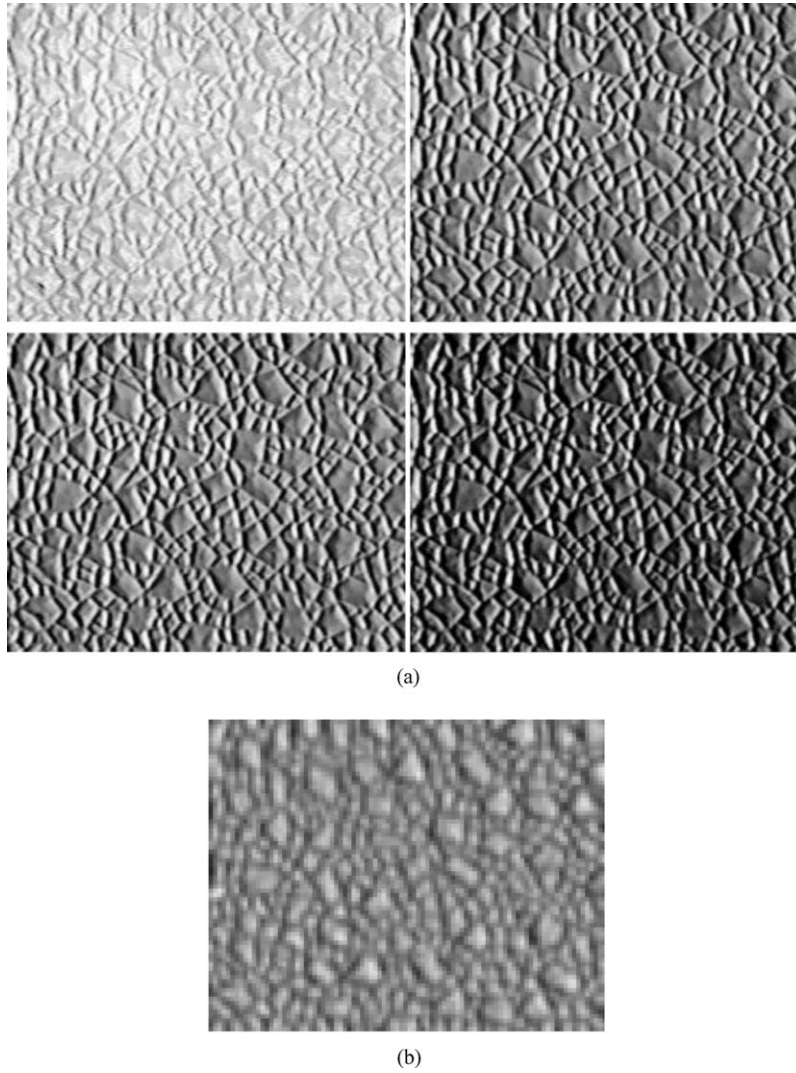

(a)

(b)

Figure 17. (a) Four calibrated gray-scale images of a rough plastic sample from the CUReT database. (b) A recovered height field from our algorithm. The height field is visualized as a gray-scale image.

Since our synthetic rendering program currently does not simulate interreflections, large shadow regions in the synthetic images remain completely dark and featureless while their counterparts in the real images have vaguely visible details. These recovered height fields can be used for displacement mapping to produce the appearance of natural rough materials.

We did another test on a rough plastic sample from the CUReT database (Fig. 17). This sample has a reasonable amount of specular reflection. Four registered input images were used for recovering the surface height field. As mentioned previously, we treat specular highlights as outliers and automatically reweight these outliers during iterative minimization. Good results have been obtained on this sample since the fea-

ture patterns are clearly visible in the recovered height field with well-defined boundaries. This indicates that our scheme for specularity is quite effective as long as the amount of specular reflection is not overly dominant. Note that there are height discontinuities at the boundaries of the feature patterns, the weight for the regularization term in the shape-from-shading formulation can be adaptively adjusted at these locations to improve the reconstruction accuracy. The details of such an adaptive scheme can be found in Liu et al. (2001).

We also did a test for a degenerate case. Three 128×128 images of a concrete sample from the CUReT database (Dana et al., 1997) were used as the input to our algorithm. They were registered together with user assistance before the experiment. From Fig. 18(a)–(c),
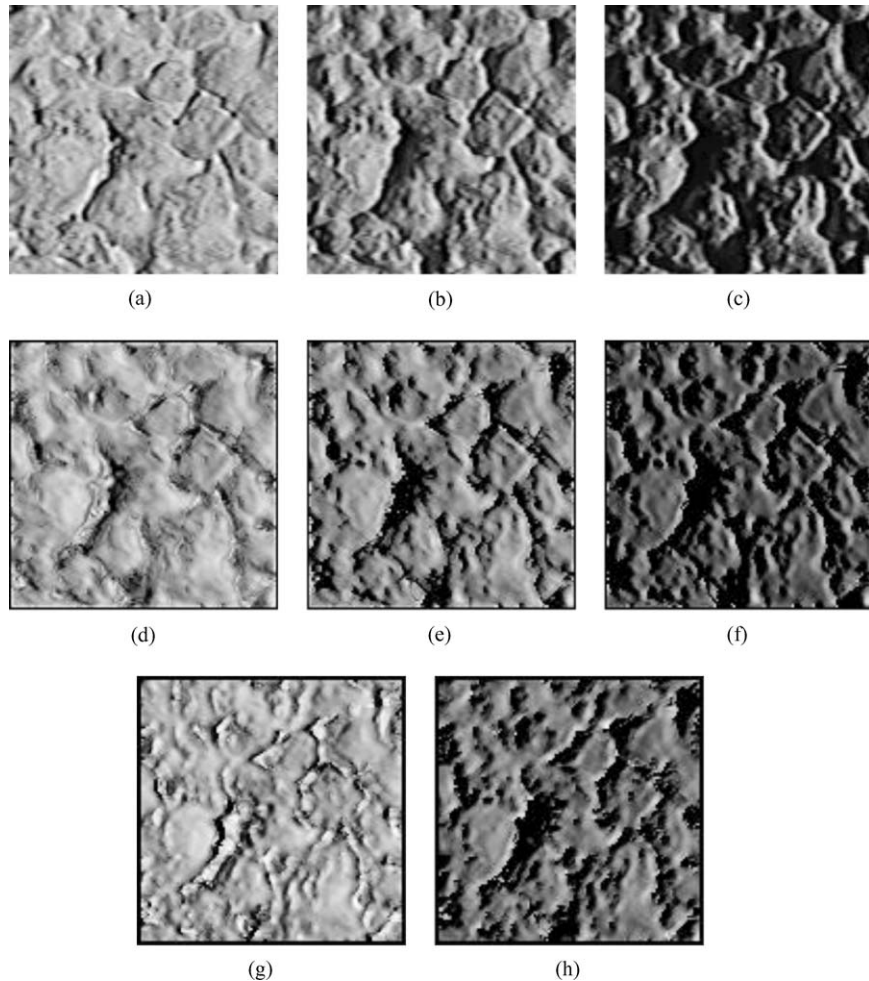

Figure 18. (a)–(c) Real images of a concrete sample from the CUReT database; (d)–(f) synthetic images of the recovered height field illuminated from original lighting directions; (g) and (h) synthetic images of the recovered height field illuminated from two novel lighting directions.

we can see they show various amount of shadow. The lighting directions of the input images are actually coplanar. Traditional photometric stereo would have problems to recover the height field. However, our algorithm obtained reasonable results since it exploits shadow constraints and a regularization term. Minimizing E_3 in Eq. (10) took 5 minutes on a Pentium III 800 MHz processor, and the iterative procedure for enforcing bounds took another half an hour. Synthetic images were generated from the recovered height field. The recovered dataset was illuminated from both original lighting directions (Fig. 18(d)–(f)) of the input images and novel lighting directions (Fig. 18(g) and (h)).

4.   Conclusions and Discussions

In this paper, we presented various shadow-based methods for recovering high-frequency surface geometric details. We introduced the concept of shadow graphs and proved that the shadow graph alone is sufficient to solve the shape-from-shadow problem from a dense set of images. We also developed a method for recovering 3D textures as well as height fields in general from both shadow and shading constraints. A constrained optimization procedure has been developed to make the results from shape-from-shading consistent with the height bounds derived from shadow constraints. Our methods are robust under noise contamination and specular reflection. The hybrid method using both shadow and shading constraints can tolerate a reasonable amount of specularity since it considers specular reflection as outliers and relies on diffuse reflection for surface recovery. As the amount of specular reflection becomes dominant such as in the case of metallic material samples, we should rely more heavily on shadow constraints since they are insensitive to specularity. Therefore, input images should be taken with a denser sampling of the illumination hemisphere in order to have more shadow constraints and more accurate height bounds from the shadow graph.

Traditional shape-from-shading (Horn and Brooks, 1986) only takes one input image and heavily relies on the regularization term. Recovered surfaces often have obvious errors. Photometric stereo (Woodham, 1989) can improve on this by using three or more images. We generalized shape-from-shading to incorporate multiple images, thus achieve better accuracy. However, it has been demonstrated that the integration of shadow constraints can further significantly improve the reconstruction accuracy using the same number of input images.

Acknowledgments

References

Belhumeur, P. and Kriegman, D. 1998. What is the set of images of an object under all possible illumination conditions? *Int. Journal Comp. Vision*, 28(3):1–16.

Belhumeur, P., Kriegman, D., and Yuille, A. 1997. The bas-relief ambiguity. In *IEEE Conf. on Comp. Vision and Patt. Recog.*, pp. 1040–1046.

Branca, A., Attolico, G., and Distante, A. 2002. Cast shadow removing in foreground segmentation. In *Intl. Conf. on Patt. Recog.*, pp. 214–217.

Dana, K.J., van Ginneken, B., Nayar, S.K., and Koenderink, J.J. 1999. Reflectance and texture of real world surfaces. *ACM Transactions on Graphics*, 18(1):1–34.

Dana, K.J. and Nayar, S.K. 1998. Histogram model for 3d textures. In *Proc. of IEEE Conf. on Comp. Vision and Patt. Recog.*

Dana, K.J. and Nayar, S.K. 1999. Correlation model for 3d textures. In *Intl. Conf. Computer Vision*.

Dana, K.J., van Ginneken, B., Nayar, S.K., and Koenderink, J.J. 1997. Reflectance and texture of real-world surfaces. In *Proceedings of CVPR*, pp. 151–157.

Daum, M. and Dudek, G. 1998. On 3-d surface reconstruction using shape from shadows. In *IEEE Conf. on Comp. Vision and Patt. Recog.*, pp. 461–468.

Dong, J. and Chantler, M. 2002. Capture and synthesis of 3d surface texture. In *The 2nd international workshop on texture analysis and synthesis*, pp. 41–45.

Dupuis, P. and Oliensis, J. 1994. Shape from shading: Provably convergent algorithms and uniqueness results. In *Computer Vision-ECCV 94*, pp. 259–268.

Fsqp software. http://gachinese.com/aemdesign/FSQPframe.htm. Originally developed at the Institute for Systems Research, University of Maryland.

Georghiades, A., Belhumeur, P., and Kriegman, D. 1999. Illumination-based image synthesis: Creating novel images of human faces under differing pose and lighting. In *IEEE Workshop on Multi-View Modeling and Analysis of Visual Scenes*, pp. 47–54.

Hampel, F.R., Rousseeuw, P.J., Ronchetti, E.M., and Stahel, W.A. 1986. *Robust Statistics*. John Wiley & Sons, New York.

Hassin, R. and Rubinstein, S. 1994. Approximations for the maximum acyclic subgraph problem. *Information Processing Letters*, 51:133–140.

Hatzitheodorou, M. 1989. The derivation of 3-d surface shape from shadows. In *Proc. Image Understanding Workshop*, pp. 1012–1020.

Horn, B.K.P. and Brooks, M.J. 1986. The variational approach to shape from shading. *Computer Vision, Graphics & Image Processing*, 33:174–208.

Ikeuchi, K. 1981. Determining surface orientations of specular surfaces by using the photometric stereo method. *IEEE Trans. Patt. Anal. Mach. Intel.*, 3(6):661–669.

Jacobs, D. 1997. Linear fitting with missing data: Applications to structure from motion and characterizing intensity images. In *IEEE Conf. on Comp. Vision and Patt. Recog.*, pp. 206–212.

Karp, R.M. 1972. Reducibility among combinatorial problems. In *Complexity of Computer Computations*, R.E. Miller and J.W. Thatcher (Eds.), Plenum: New York, pp. 85–103.

Kender, J. and Smith, E. 1987. Shape from darkness. In *Int. Conf. on Computer Vision*, pp. 539–546.

Kriegman, D.J. and Belhumeur, P.N. 1998. What shadows reveal about object structure. In *Computer Vision-ECCV 98*.

Langer, M., Dudek, G., and Zucker, S.W. 1995. Space occupancy using multiple shadowimages. In *Proceedings IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 390–396.

Leclerc, Y.G. and Bobick, A.F. 1991. The direct computation of height from shading. In *Proc. of IEEE Conf. on Comp. Vision and Patt. Recog.*, pp. 552–558.

Leung, T. and Malik, J. 1999. Recognizing surfaces using three dimensional textons. In *Intl. Conf. Computer Vision*.

Liu, X., Yu, Y., and Shum, H.-Y. 2001. Synthesizing bidirectional texture functions for real-world surfaces. In *Proceedings of SIGGRAPH*, pp. 97–106.

Medioni, G. 1983. Obtaining 3-d from shadows in aerial images. In *IEEE Conf. on Comp. Vision and Patt. Recog.*, pp. 73–76.

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

Nayar, S.K., Ikeuchi, K., and Kanade, T. 1991. Shape from inter-reflections. *International Journal of Computer Vision*, 6(3):2–11.

Prati, A., Mikic, I., Trivedi, M.M., and Cucchiara, R. 2003. Detecting moving shadows: Algorithms and evaluation. *IEEE Patt. Anal. Mach. Intel.*, 25(7):918–923.

Press, W.H., Flannery, B.P., Teukolsky, S.A., and Vetterling, W.T. 1988. *Numerical Recipes in C*. Cambridge Univ. Press, New York.

Shafer, S.A. 1985. *Shadows and Silhouettes in Computer Vision*. Kluwer Academic Publishers.

Shi, J. and Malik, J. 2000. Normalized cuts and image segmentation. *IEEE Trans. Pat. Anal. Mach. Intell.*, 22(8):888–905.

Stewart, A.J. and Langer, M.S. 1997. Towards accurate recovery of shape from shading under diffuse lighting. *IEEE Patt. Anal. Mach. Intel.*, 19(9):1020–1025.

Terzopoulos, D. 1988. The computation of visible-surface representations. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 10(4):417–438

Van Gool, L., Vandermeulen, D., Kalberer, G., Tuytelaars, T., and Zalesny, A. 2002. Modeling shapes and textures from images: New frontiers. In *Symp. 3D Data Processing Visualization and Transmission*, pp. 286–294.

Woodham, R.J. 1989. Photometric method for determining surface orientation from multiple images. In *Shape from Shading*, B.K.P. Horn and M.J. Brooks (Eds.), MIT Press, pp. 513–532.

Yu, Y. and Chang, J.T. 2002. Shadow graphs and surface reconstruction. In *7th European Conference on Computer Vision*, vol. II, pp. 31–45.