Parametric Shape-from-Shading by Radial Basis Functions

Guo-Qing Wei, Member, IEEE, and Gerd Hirzinger, Senior Member, IEEE

Abstract—In this paper, we present a new method of shape from shading by using radial basis functions to parameterize the object depth. The radial basis functions are deformed by adjusting their centers, widths, and weights such that the intensity errors are minimized. The initial centers and widths are arranged hierarchically to speed up convergence and to stabilize the solution. Although the smoothness constraint is used, it can be eventually dropped out without causing instabilities in the solution. An important feature of our parametric shape-from-shading method is that it offers a unified framework for integration of multiple sensory information. We show that knowledge about surface depth and/or surface normals anywhere in the image can be easily incorporated into the shape from shading process. It is further demonstrated that even qualitative knowledge can be used in shape from shading to improve 3D reconstruction. Experimental comparisons of our method with several existing ones are made by using both synthetic and real images. Results show that our solution is more accurate than the others.

Index Terms—Shape from shading, radial basis functions, hierarchical structure, depth constraints, normals constraints, qualitative knowledge, stochastic gradient method.

1 INTRODUCTION

Despite recent advances in depth estimation, e.g., range sensors or stereo vision, shape from shading remains an important alternative means for estimating shape. This is because, range sensors have limited acting range, and stereo relies heavily on textured regions. Shape from shading complements such methods just in these aspects.

Since the mathematical formulation and the numerical solution of the shape from shading (SFS) problem by Horn [10], there have been a lot of recent developments [12]. Local methods [26], [17], [6], recover the surface normal at a point independently of the other points, based only on local intensity values and a local geometrical model (e.g., sphere). Semi-local methods either propagate surface depths from given equal-height contours [3], [15] or propagate surface gradients from singular points in down-hill directions toward the light source [5]. Global methods, employing either variational reformulations [13], [1], [2], [11], [40], or parameter optimizations [16], [34], [18], [38], requires no explicit knowledge about the object surface; and it has been found that global methods are more robust than local or semi-local ones [39]. The following issues have been the central points of research on global shape from shading methods in the literature.

1.1 Smoothness

Many shape-from-shading methods require an explicit smoothness constraint to be included in the cost function or functional to find a physically meaningful solution. The conventional smoothness constraint is a quadratic function of the gradient derivatives [13], [1], [2], [11]. Since it tends to over-smooth the reconstructed surface, Horn [11] suggested to reduce the smoothness weight gradually (from 1.0) to a lower limit during the iteration. Choosing the optimal manner for reducing the smoothness weight, especially, determining the optimal lower limit turns out to be a non-easy task. Zheng and Chellapa [40] implemented a smoothness constraint by requiring the intensity gradients of the given and the reconstructed images to be equal. Their smoothness weight was fixed at 1.0 during the iteration. Jones and Taylor [14] used superimpositions of Gaussians at different (fixed) scales to recover the object surface. Wei and Hirzinger [38] used a multilayer perceptron to parameterize the object surface. Both the methods do not use explicitly a smoothness constraint.

1.2 Integrability

When the surface depth and surface gradients are treated as independent variables, as in many variational approaches of SFS, e.g., [11], [40], [9], the integrability constraint has to be imposed, which relates the depth and gradients in a weaker manner. Although the integrability weights can be fixed during iteration, it was found that the value of the integrability weight affects the solution, too [38]. The integrability constraint could be avoided, if one solves for the depth directly without using surface gradients as intermediate variables. For variational approaches, however, this would result in a fourth, (or higher) order partial differential equation, whose solution may encounter problems at the image boundary. This problem does not occur in parameter-optimization approaches [16], [18], [14], [38].

1.3 Convergence Rate

The finite difference discretization of the Euler equations in variational approaches (e.g., [11]) and the use of local sur-
face representations in optimization approaches (e.g., the finite element approach [18]) often result in a low convergence rate since only local image pixels interact. (The propagation of low frequency surface components is thus inefficient, although the intensity error may go down very quickly.) Terzopoulos [35] proposed to use multigrid methods to speed up convergence by propagating results in both the coarse-to-fine and the fine-to-coarse directions. Szeliski [34] made an advance in accelerating convergence by using hierarchical basis functions to approximate both the depth and the gradients in Horn’s objective function [11]. (The smoothness and the integrability constraints were still used in [34].) Using global surface representations such as the sigmoid function network [38] can propagate very quickly the low frequency components, but it is not flexible enough to model high frequency surface components.

In this paper, we present a parametric shape-from-shading method, which uses hierarchical (both global and local) radial-basis-functions (Gaussians) to represent the object depth. The method is an extension of our previous work [38], and bears close resemblance to several other publications, in the sense that the representation is hierarchical [34], [14], and bears close resemblance to several other publications, in the sense that the representation is hierarchical [34], [14]. But the following differences are obvious. Contrary to [34], [18], [14], the shapes of the basis functions in our method are not fixed, but subject to deformations according to the data: Besides the Gaussians’ weights, the centers and widths are also to be adjusted in the continuous domain. This use of variable basis functions makes the reconstruction results very insensitive to the number of basis functions chosen. In fact, we let the number of basis functions adaptive to the errors of the objective functions through an incremental scheme. Although a smoothness constraint is used, it can be eventually dropped out for further iterations without deteriorating the result. Another salient feature of our work is that we show how to integrate a priori knowledge about depth and surface normals into the shape from shading process. Exploring ways of applying prior knowledge in shape from shading is an important issue and has not been given much attention in the literature. We demonstrate that it is even possible in shape from shading to use qualitative knowledge to improve reconstruction results. Depending on how much the prior knowledge is with respect to shading information, our method can be either viewed as applying prior knowledge to shape from shading or regarded as using shading to aid surface reconstruction/interpolation. In this sense, we have unified the two frameworks of shape from shading and 3D reconstruction/interpolation.

The paper is organized as follows. In Section 2, we describe our shape from shading method. In Section 3, we show how to apply a priori knowledge in shape from shading estimation. In Section 4, we illustrate some computational considerations. In Section 5, we compare our methods with several existing ones. Finally we give conclusions in Section 6.

2 Depth from Shading

2.1 Reflectance Map

We assume a Lambertian reflectance model. Hybrid reflectance models can be processed to produce a pure Lambertian map by using existing methods, e.g., [32], [37]. Suppose that the surface of an object is represented by \( z = z(x, y) \) in the camera coordinate system \( x : y : z \), with the \( x\)-\( y \) plane coinciding with the image plane, and the \( z \) axis coinciding with the optical axis of the camera. Assume orthographic projection for image formation (extensions to perspective projections can be made similarly to [19]). Under these assumptions, the image intensity at position \((x, y)\) of the image plane can be computed as [10]

\[
I(x, y) = \eta \mathbf{n} \cdot \vec{L} = R(p, q) = \eta \frac{-p l_1 - q l_2 + l_3}{\sqrt{p^2 + q^2 + 1}}
\]

where \( \eta \) is the composite albedo, \( \mathbf{n} \) is the surface normal at \([x, y, z(x, y)]\),

\[
\mathbf{n} = (n_1, n_2, n_3)^T = \begin{pmatrix}
  -p \\
  -q \\
  1
\end{pmatrix} / \sqrt{p^2 + q^2 + 1}
\]

\[
p = \frac{\partial z}{\partial x}, \quad q = \frac{\partial z}{\partial y}
\]

and \( p \) and \( q \) are the surface gradients at \((x, y)\); \( \vec{L} = (l_1, l_2, l_3) \) is the illuminant direction. Equation (1) is called the image irradiance equation, which is a first order nonlinear partial differential equation on \( z \); \( R(p, q) \) is often referred to as the reflectance map.

Suppose in the following that the parameters \( \eta \) and \( \vec{L} \) are known. The shape from shading problem can thus be stated as the determination of \( p(x, y), q(x, y), \) and \( z(x, y) \) from a given image \( I(x, y) \).

2.2 Depth Representation

Our idea of parametric shape-from-shading is to parameterize the depth function \( z(x, y) \). Since no a priori knowledge about the object shape is assumed available, the parameterization should be flexible enough to approximate any surfaces. There have been results that radial-basis-function (RBF) networks are capable of universal approximations [25]. The RBF networks were mainly used for interpolation/approximation purposes, where 3D data at certain positions have already been given [29], [28], [8], while the use of RBFS in shape from shading is much more underconstrained, since only intensity data are available; the purpose is to reconstruct 3D data. There are several types of radial basis functions used in interpolations [29]. We choose the Gaussian radial basis function since it is of finite support and it is separable in the \( x \) and \( y \) directions.

With Gaussian radial basis functions, the surface depth \( z(x, y) \) can be modeled as

\[
z(x, y) = \sum_{k=1}^{N} w_k \phi(x, t_k, s_k) = w_o
\]
where \( N \) is the number of Gaussian functions, and \( x = (x, y) \); parameters \( t_k = (l_{x,k}, l_{y,k}) \), \( s_k = (\sigma_{x,k}, \sigma_{y,k}) \), and \( w_k \) are, respectively, the center, width (size), and the weight of the \( k \)th Gaussian function:

\[
\phi(x, t_k, s_k) = e^{-\frac{(x-t_{x,k})^2 + (y-t_{y,k})^2}{\sigma_{x,k}^2 + \sigma_{y,k}^2}} \tag{5}
\]

Each Gaussian has a receptive field \([x \in t_{x,k} \pm 3\sigma_{x,k}, y \in t_{y,k} \pm 3\sigma_{y,k}]\), beyond which its contribution can be ignored. Therefore the Gaussian RBFs are locally tuned.

### 2.3 Depth from Shading

If we use the vector \( \mathbf{W} = \{ t_k, s_k, w_k \} \) to represent all the Gaussian parameters in the specification of depth \( z(x,y) \), then the surface gradients can be analytically computed as

\[
p(x,y) = \frac{\partial z(x,y, \mathbf{W})}{\partial x}, \quad q(x,y) = \frac{\partial z(x,y, \mathbf{W})}{\partial y} \tag{6}
\]

which are functions of the same unknown parameters \( \mathbf{W} \). Substituting (6) into (1) results in an equation on \( \mathbf{W} \):

\[
R(z_i(x,y, \mathbf{W}), z_j(x,y, \mathbf{W})) = \eta \frac{z_i(x,y, \mathbf{W}) - z_j(x,y, \mathbf{W})}{\sqrt{(z_i(x,y, \mathbf{W}) - z_j(x,y, \mathbf{W}))^2 + 1}} \tag{7}
\]

The shape from shading problem is then to seek an \( \mathbf{W} \) which minimizes the following cost function \( E_{IS} \):

\[
E_{IS} = E_I + \lambda E_S = \sum_{i \in D_I} E_{IS,i}
\]

\[
= \sum_{i \in D_I} \left[ l_i - R(z_i(x,y, \mathbf{W}), z_j(x,y, \mathbf{W})) \right]^2 + 2S_1(x,y)z_{xj}(x,y, W) + 2S_2(x,y)z_{yj}(x,y, \mathbf{W}) + S_3(x,y)\frac{z_{xy}(x,y, W)}{1 + \sqrt{(z_i(x,y, \mathbf{W}) - z_j(x,y, \mathbf{W}))^2 + 1}} \tag{8}
\]

where \( E_I \) and \( E_S \) are the intensity error and the smoothness error, respectively; \( \lambda \) is a global smoothness weight; \( D_I \) is the index set of all the image points; \( I_i \) is the given intensity value at the \( i \)th pixel located at \( (x, y) \); and \( \{ S_j(x,y), j = 1, 2, 3 \} \) are local smoothness weights:

\[
S_1(x,y) = \left[ 1 - \frac{1}{I_x(x,y)} \right]^2 \tag{9}
\]

\[
S_2(x,y) = \left[ 1 - \frac{\sqrt{2}}{2} \frac{I_y(x,y) + I_x(x,y)}{I_x(x,y)} \right]^2 \tag{10}
\]

\[
S_3(x,y) = \left[ 1 - \frac{1}{I_y(x,y)} \right]^2 \tag{11}
\]

which are chosen to be inversely proportional to the intensity gradients along the \( x- \), \( y- \) diagonal, and \( y- \) directions, respectively. This choice is intuitive, but it conforms well with the fact that smoother images should have been produced by smoother surfaces, assuming no albedo variations. (Here we have normalized both the input image and the reflectance map by the albedo, so that both \( I_x \) and \( I_y \) are less than one.)

The introduction of the local smoothness weights in (8) is analogous to the use of controlled-continuity in surface interpolation [36], [21], where the local smoothness weights can yet be composed directly of the surface to be reconstructed; that is, \( S_j(x,y) \) may be functions of \( z_{xx}, z_{yy}, z_{yx}, z_{xy} \) and \( z_{yy} \) as in [21]. We found, however, that this does not yield a reasonable 3D reconstruction in the shape from shading problem, probably because shape from shading is much more underconstrained than 3D interpolation/reconstruction. We also tried to replace \( E_S \) in (8) by Zheng and Chellapa’s smoothness term [40]. But since their constraint is weaker than the quadratic one, its use in our formulation produces a surface which conforms only locally well, but is globally too ragged. (Zheng and Chellapa’s smoothness constraint provides two constraint-equations on the three variables \( z_{xx}, z_{yy}, \) and \( z_{xy} \), while the quadratic smoothness constraint in (8) provide three constraints on them; here the null space is \( z_{xy} = 0, z_{yy} = 0 \) and \( z_{yy} = 0 \).) To summarize, the quadratic smoothness constraint with the local weights determined by intensity gradients has been found to give the most reasonable 3D reconstructions.

Now, to minimize (8), there are many optimization approaches that can be used [30]. Gradient descent, when applied directly to (8), is too slow in convergence; whereas the conjugate gradient method is inefficient here since it needs many function evaluations where the number of Gaussians \( N \) is usually very large. We choose the stochastic gradient method, which adjust the parameters of the Gaussians pixel by pixel cyclically, based on the gradients of \( E_{IS,j} \). That is, for the \( i \)th pixel, the parameters are updated according to

\[
\Delta \mathbf{W}^{(n)} = -\beta \frac{\partial E_{IS}}{\partial \mathbf{W}} + \alpha \Delta \mathbf{W}^{(n-1)} \tag{12}
\]

where \( \Delta \mathbf{W}^{(n)} \) represents the change of \( \mathbf{W} \) for the current pixel, \( n \) indexes the number of iterations; \( \Delta \mathbf{W}^{(n-1)} \) represents the update for the previous pixel; \( \beta \) is the learning rate, and \( \alpha \) is a momentum term [31] used to damp possible oscillations. Typical values for \( \beta \) and \( \alpha \) are \( \beta = 0.3 \) and \( \alpha = 0.6 \) in our experiments. Since the Gaussians are of finite support, there are usually only a small portion among the \( N \) Gaussians which are active for a pixel. This makes the use of stochastic gradient descent more economic since only the active Gaussians need to be updated. During the iteration, the learning rate \( \beta \) should be decreased whenever the total error \( E_{IS} \) shows any oscillations. We reduce it by 30 percent of its current value. Convergence is reached when the change of \( E_{IS} \) becomes insignificant.

Notice that in our solution above we do not need any specific programming considerations at the image boundary or boundaries of regions beyond which the image intensities are either not available or not reliable. (We only need to exclude the unreliable points from the index set \( D_I \).) In variational approaches, however, natural boundary conditions should be imposed on such boundaries, since the Euler equations cannot be applied there. To impose the
natural boundary conditions, one has to identify the boundary pixels and to consider all possible boundary shapes near a boundary pixel explicitly in the computer program (e.g., see [11] for vertical and horizontal boundaries). (Note that the shape of such boundaries may be irregular.) Note also that natural boundary conditions serve only as a set of necessary conditions for functional minimization [4, p. 208]. It does not necessarily ensure the 3D reconstruction to be physically correct.

### 3 Applying Prior Knowledge

When there are no boundary conditions available at the image boundary (i.e., free boundary), or when there are depth discontinuities or shadows in the image, the recovered surface by intensity information alone may be inaccurate near these points. Although no prior knowledge about the object shape could be really expected at the image boundary, edges within the image, resulting from depth discontinuities, for instance, may be used by stereo vision to provide depth or surface normal information. Here, we consider a general case in which surface depths or normals of random values was found to be inefficient both in computations and in reconstructing a reasonable surface. Here, we would like to mention that Shao, Chellapa, and Simchony [33] also proposed to reconstruct a 3D surface by combining information from different sources, e.g., a needle map obtained by photometric stereo and a depth map obtained from geometric stereo. Their formulation, though extendable to include shading information, was based on the variational formulation, and therefore, suffers from the same problems mentioned in the Introduction. Besides, no qualitative constraints were considered.

#### 4 Implementations

4.1 Computational Considerations

4.1.1 Hierarchical Gaussians

In interpolation problems, the starting values for \{t_i, s_j, w_k\} in the iteration can be either obtained by analyzing the given 3D data [8], [22], or just set to random values. In shape from shading, 3D data are not available; and the use of random values was found to be inefficient both in computations and in reconstructing a reasonable surface. Here...
we employ a heuristic approach. We let the initial Gaussians’ centers regularly distributed in the image plane and let the initial widths cover different scales. This is similar in style to the use of hierarchical basis functions [34] or scale-space [14]. However, the centers and widths (scales) in our case are subject to continuous adjustment in accordance to the image data. This makes the number of scales very small in comparison to that in [14].

Suppose $2^M \times 2^M$ is the image size. We set the maximum number of hierarchy to be $H$, e.g., $H = M - 1$. The number of Gaussians in hierarchy $h$ is chosen to be $2^k \times 2^k$, $h = 1, 2, ..., H$. Within a hierarchy, we set the initial positions $\{t\}$ to be equally spaced in the image plane, and the initial widths to be equal to the spacing; that is, at hierarchy $h$, the initial spacing and width are $2^{h-1}/2^h$ in both the $x$ and $y$ directions. The initial weights of all the Gaussians are set to zero.

4.1.2 Incremental Scheme

It is computationally inefficient to use all the Gaussians at all the hierarchies to start the iteration. It was also found that Gaussians at higher hierarchies are adjusted more strongly than those at lower hierarchies, lessening the contribution of coarser Gaussians to the reduction of error functions. This, together with the local support of Gaussians, inspires us to use an incremental scheme to build the object surface. First, the Gaussians at the lowest hierarchy, $h = 1$, are used to minimize the errors of the objective functions. After the convergence at the current hierarchy, we add Gaussians at the higher hierarchy. For each new Gaussian to be added, e.g., the $i$th one, we check the average intensity error $e_I$, depth error $e_d$, and normal error $e_n$ within its main receptive field defined as $[x \in t_{ij} \pm 1.5\sigma_x, y \in l_{ij} \pm 1.5\sigma_y]$. If all the errors $e_I$, $e_d$, and $e_n$ are below the respective thresholds, the $i$th Gaussian will not be created. When new Gaussians are generated, we do not freeze the old ones; that is, the old Gaussians are still subject to further adjustment. This avoids possible local minima according to our experience. The incremental scheme spares lots of unnecessary Gaussians (up to 50 percent).

4.1.3 Multiresolution

To further speed up computations, we also use a multiresolution scheme through subsampling. In our parametric approach of shape from shading, the use of multiresolution is just to reduce the number of pixels used in the iteration. The coordinates of the pixels do not change, since it is the same parametric function representing the depth in the finest resolution that is used to minimize the errors of the objective functions at different resolutions. So there is no need to use explicit interpolation of depth in the transition from a lower resolution to the higher one. The coarsest resolution we choose is $32 \times 32$. For the depth or normal constraints, no subsampling is used since they are usually few in number. The multiresolution scheme works together with the incremental scheme. At each resolution, the highest hierarchy allowed is chosen to be consisting of Gaussians whose initial spacing is two times that between neighboring pixels at that resolution; for instance, at the $32 \times 32$ image resolution, the highest hierarchy is composed of Gaussians which are initially spaced in the form of $16 \times 16$.

4.2 Computational Steps

In the following, we outline the computational steps in the implementation of our algorithm. $E_{\text{total}}$ is used to represent the total sums of all the errors: intensity error, depth error, normal error, and smoothness error.

Step 1) normalize the image intensity to the range $[0, 1]$; subsample the image to $L$ levels. Compute $I_x$, $I_y$. Set $\lambda = 0.1$. Set the current resolution level $l$ to $l = 1$. Set the number of iterations $n$ to 0. Set the starting hierarchy to $h = 1$, and set the initial RBFs at $h = 1$.

Step 2) for each $i \in D_d(l)$, $k \in D_d$ and $j \in D_d$: Adjust the RBFs using the stochastic gradient method for one cycle; $n \leftarrow n + 1$.

Step 3) check whether to reduce the learning rate $\beta$, depending on whether $E_{\text{total}}$ oscillates.

Step 4) if $\frac{E_{\text{total}}^{(n)} - E_{\text{total}}^{(n-1)}}{E_{\text{total}}^{(n)}} > 0.1$ percent, go to Step 2;

else if the current RBFs have been in the highest hierarchy set for the current resolution $l$

\[
\begin{cases}
\text{if } l = L, \text{go to Step 5}; \\
\text{if } l \neq L, \text{then } n \leftarrow l + 1, \text{go to Step 2};
\end{cases}
\]

else generate new RBFs at the higher hierarchy at positions where $e_I > 0.5 \cdot \text{rms}(E_d)$, or $e_d > 0.5 \cdot \text{rms}(E_d)$, or $e_n > 0.5 \cdot \text{rms}(E_n); h \leftarrow h + 1$; go to Step 2;

Step 5) when necessary, set $\lambda$ to 0, and repeat Step 2 until convergence is reached; otherwise, stop.

In Step 2, one cycle means a complete pass of all the pixels in $D_d(l)$, $D_d$, and $D_d$; here $D_d(l)$ signifies the dependence of the number of image points on the resolution level $l$. In Step 4, the generation of a new Gaussian is made adaptive to both the local errors $e_I$, $e_d$, $e_n$ within the major receptive field and the current global rms error, where $\text{rms}(E)$ indicates the rms error of $E$. (The requirement that local errors be larger than 50 percent of the global errors is purely empirical.) The global smoothness weight $\lambda$ is empirically fixed at 0.1. After convergence, it could be completely removed for further iterations if there were no image noises. In Step 3, $\beta$ ceases to decrease when it reaches $\beta = \beta_{\text{min}}$; here $\beta_{\text{min}} = 0.01$.

5 Experiments

In this section, we first exemplify our parametric shape-from-shading method. Then we compare our method with several recently published ones. To avoid an exhaustive listing of all the methods, we only include those which were not examined (or not properly implemented) in [39]. (A comprehensive survey and implementation of many shape-from-shading algorithms has been reported in [39]. The implemented methods there include [2], [5], [16], [17], [18], and [40], among others.) The methods considered in our paper are Horn’s method [11], Zheng and Challappa’s method [40], and the Wavelet method (Hsieh et al. [9]).

1. In [39], the implementation of Zheng and Challappa’s method [40] does not conform with the authors’ original implementation at the image boundary. The results reported in [39] are much poorer than those in the authors’ original publications [40].
be true to the original implementation, our implementations are based on the original C or Fortran codes provided by the authors (except for Horn’s method). The comparison will be performed in two aspects:

1) shape from pure shading, and
2) incorporating a priori knowledge.

All the images we use had appeared in previous shape-from-shading publications, and are accessible to the public. Among the test images, two are synthetic (Sombrero, Mozart) and four are real (David, Agrippa, left; Pepper, and Agrippa, right). The albedos and illuminant directions of these images are listed in Table 1. The image sizes we choose to work with are all $64 \times 64$.

<table>
<thead>
<tr>
<th>Image</th>
<th>Albedo $\eta$</th>
<th>Illuminant direction (deg)</th>
<th>Slant (deg)</th>
<th>Tilt (deg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>David</td>
<td>190.0</td>
<td>135.0, 50.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mozart</td>
<td>255.0</td>
<td>135.0, 45.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pepper</td>
<td>245.0</td>
<td>40.0, 45.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agrippa left</td>
<td>190.0</td>
<td>135.0, 50.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sombrero</td>
<td>255.0</td>
<td>0.0, 45.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agrippa right</td>
<td>190.0</td>
<td>45.0, 45.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### 5.1 An Example of Our Method

Fig. 1a shows a real image of the David statue. The image contains both relatively flat area (the cheek) and highly structured area (the eye and nose). It is best suited for demonstrating how Gaussians have been generated and deformed to adapt to the image structure. We use two levels of image resolutions, which correspond to the image sizes of $32 \times 32$ and $64 \times 64$, respectively. Five hierarchies of Gaussians are used, which are distributed over the two image resolutions. At the $32 \times 32$ image resolution, hierarchies of Gaussians corresponding to the initial positioning in the forms of $2 \times 2$, $4 \times 4$, $8 \times 8$, and $16 \times 16$ are generated; while at the $64 \times 64$ image resolution, only the hierarchy containing Gaussians initially spaced in the form of $32 \times 32$ is generated (see Section 4.1). As has been mentioned, the initial Gaussians at each hierarchy are equally spaced in the original $64 \times 64$ image plane. Using the incremental scheme described in Section 4, a total of 691 Gaussians have been generated after 455 cycles of iterations when convergence is reached. (This takes about 10 minutes in an Indigo2 Silicon Graphics workstation; the running time could be reduced by about a factor of five with a looser convergence criterion.) Fig. 1b shows the reconstructed image, and Fig. 1c depicts the distribution of the Gaussians after convergence. It can be easily seen that at smoother area, the Gaussians are sparsely distributed; while at structured area, the Gaussians are densely distributed. The Gaussians have been deformed both in positions and in shapes. The recovered 3D shape illuminated from another source direction is shown in Fig. 1d.

Throughout the examples shown in this paper, we have kept the smoothness weight $\lambda$ fixed at 0.1. We found that after convergence, the smoothness term could be entirely removed for further iterations without destroying the result, seemingly due to the presence of global Gaussians which stabilize the solution; that is, the result is insensitive to the lower limit of the smoothness weight. This is not the case with variational approaches: In Horn’s method [11] (and in others), removing the smoothness term (i.e., setting $\lambda = 0$) may cause raggedness in the recovered surface, even though the intensity error remains at a very small level (results not shown here).

### 5.2 Shape From Pure Shading: A Comparison

First, a few words about the implementation of Horn’s, Zheng and Chellapa’s, and the Wavelet methods are needed, since an improper implementation may lead to poor results or even disconvergence. (These details could not be found in their original papers.) In Horn’s method, we use the Jacobi update for both $p$, $q$, and $s$. The smoothness weight is reduced from 1.0 to 0.05 during iteration: Whenever the change of the total system error is less than 0.0001 percent, $\lambda$ is reduced by 10 percent of the current value. In Zheng’s method, the update for $p$, $q$, and $s$ is parallel, all using the old values; while at the image boundary, the updates use new values (according to their Fortran code). The Wavelet method by Hsieh et al. [9] is in fact the use of finite-differences of a larger support (nine pixels) to discretize the Euler equations (i.e., to approximate $p_x$, $p_{xx}$, $p_{xxx}$, and $z_{xxxx}$ etc. in their Euler equations). The updates for $p$, $q$, and $z$ are all sequential,

2. An erratum in [9]: In equation (22), the indices $i + k$ and $j + k$ should be corrected as $i - k, j - k$, as was indicated by their C-code.
TABLE 2  
THE ACCURACY OF DIFFERENT METHODS UNDER DIFFERENT BOUNDARY CONDITIONS FOR THE SOMBRERO IMAGE

<table>
<thead>
<tr>
<th>boundary condition</th>
<th>methods</th>
<th>depth error (rms, std, max)</th>
<th>normal error (rms, std, max)</th>
</tr>
</thead>
<tbody>
<tr>
<td>free boundary</td>
<td>Ours</td>
<td>-2.21, 4.41</td>
<td>9.65, 5.99, 79.5</td>
</tr>
<tr>
<td></td>
<td>Horn</td>
<td>-3.19, 5.38</td>
<td>12.7, 9.35, 75.7</td>
</tr>
<tr>
<td></td>
<td>Zheng</td>
<td>-3.17, 4.95</td>
<td>13.7, 10.3, 72.8</td>
</tr>
<tr>
<td></td>
<td>Wavelet</td>
<td>-1.76, 5.37</td>
<td>16.1, 11.9, 74.5</td>
</tr>
<tr>
<td>boundary normal</td>
<td>Ours</td>
<td>-1.58, 3.22</td>
<td>6.15, 4.52, 79.8</td>
</tr>
<tr>
<td></td>
<td>Horn</td>
<td>-2.98, 5.09</td>
<td>11.3, 8.98, 75.7</td>
</tr>
<tr>
<td></td>
<td>Zheng</td>
<td>-2.97, 4.93</td>
<td>11.1, 8.92, 72.6</td>
</tr>
<tr>
<td></td>
<td>Wavelet</td>
<td>-1.89, 6.02</td>
<td>16.7, 14.4, 74.5</td>
</tr>
<tr>
<td>boundary depth</td>
<td>Ours</td>
<td>0.35, 0.37, 2.49</td>
<td>4.05, 3.72, 79.9</td>
</tr>
<tr>
<td></td>
<td>Horn</td>
<td>0.94, 0.94, 3.73</td>
<td>7.66, 6.05, 75.5</td>
</tr>
<tr>
<td></td>
<td>Zheng</td>
<td>0.94, 0.98, 3.72</td>
<td>8.28, 6.38, 72.3</td>
</tr>
<tr>
<td></td>
<td>Wavelet</td>
<td>1.61, 1.46, 7.27</td>
<td>9.64, 7.52, 74.5</td>
</tr>
<tr>
<td>boundary normal &amp; depth</td>
<td>Ours</td>
<td>0.30, 0.35, 2.40</td>
<td>3.36, 3.69, 78.5</td>
</tr>
<tr>
<td></td>
<td>Horn</td>
<td>0.71, 0.72, 4.17</td>
<td>6.08, 5.47, 75.6</td>
</tr>
<tr>
<td></td>
<td>Zheng</td>
<td>0.65, 0.68, 3.65</td>
<td>6.62, 5.53, 72.2</td>
</tr>
<tr>
<td></td>
<td>Wavelet</td>
<td>2.50, 1.91, 9.04</td>
<td>12.8, 10.6, 74.5</td>
</tr>
</tbody>
</table>

Fig. 2. The Mozart image. (a) The 3D ground truth. (b) The input image. (c) The reconstructed surfaces by our method. (d) Horn’s method. (e) Zheng and Chellapa’s method. (f) The Wavelet method.
using the new values (Gauss-Seidel). Because of the finite differences of large supports, variables outside the image boundary are simply copied from those at the image boundary. (This is theoretically not justified since such a copying scheme contradicts the corresponding natural boundary conditions, which are, due to the involvement of high order derivatives, though difficult to implement.) In both Horn’s and Zheng’s methods, intensities are normalized to the range $[0, 1]$; while in the Wavelet method, no normalization is used (according to their $C$ code); this is equivalent to using smaller smoothness and integrability weights. Multiresolution is used in all the three methods. Convergence is reached when the system error does not change by more than 0.0001 percent. For the Wavelet method, the system error may increase after it has decreased to some level. Convergence is then chosen to be at the point of minimum system error.

1) The Mozart Image: Fig. 2a plots the range data of a Mozart surface. (We choose a viewing angle with which the fine structure can be seen more clearly). The 3D shape contains both fine structure and depth discontinuities. Fig. 2b is a synthetic image from the range data. Since the ground truth itself has defects, we shall only judge 3D reconstructions subjectively. Figs. 2c-f show, respectively, the depths recovered by our method, Horn’s method, Zheng and Chellapa’s method, and the Wavelet method. (Scales in the $z$ direction have been made equal for the four plots. This is the case for all later comparisons, since scales in the $z$-direction may alter appearances.) Subjective judgment indicates that our method, Horn’s method, and Zheng and Chellapa’s method produce qualitatively similar results; while the Wavelet method has difficulty in recovering the fine structure. By comparing
the recovered face orientations and the relative heights of the noises, the height of protrusion of the faces from the background, we can see that the result obtained by our method resembles more closely the original 3D shape. The results of Horn’s, and Zheng and Chellapa’s methods tends to be smoother than the original shape.

2) The Pepper Image: The observations in the last example are confirmed by another example, using a real Pepper image (Fig. 3a). Shown in Figs. 3b-e are, respectively, the reconstructed depths by the four methods, as viewed from South-West. It can be seen that the peppers recovered Horn’s, Zheng and Chellapa’s, and the Wavelet methods are too flat, while our method reconstructs a pepper which looks more natural. The result can be seen more clearly if one plots the surface normals projected on the image plane.

3) The David Image: This example illustrates the behaviour of the four methods at image boundaries when there is no enough intensity structure. The input image has been shown in Fig. 1a. Figs. 4a-d show, respectively, the recovered surfaces by the four methods. It can be seen that the recovered shape of our method conforms better with the human perception, while the other methods are subject to more errors at the image boundary (lower-left corner), although the reconstructed image of each method was found to be very close to the input image.

4) The Agrippa Image: Fig. 5a shows a real image of the Agrippa stature (Agrippa left). Figs. 5b-e show the recovered surfaces of the four methods. The methods of Horn, Zheng and Chellapa, and Wavelet have the same problem as in the last example at the image boundary, while our method performs much better. Notice that the left and right cheeks recovered by Horn’s and Zheng and Chellapa’s methods are not at the same height—an apparent error due to the shadows of the forehead and of the nose, which almost isolate the right cheek. The result achieved by our method has been, to our surprise, almost the same as that obtained by the photometric stereo shape-from-shading method using two images at different illuminant directions (see Fig. 17c of [20]).

5.3 Applying Prior Knowledge
Here we use a synthetic image whose ground truth is exactly known to evaluate the performances of the four methods in the presence of various a priori knowledge. Although our method can incorporate knowledge anywhere in the image, we show here an example in which prior knowledge is given at the image boundary. Fig. 6a is the ground truth (Sombrero). The input image is illustrated in Fig. 6b. We consider the following four situations in which a priori knowledge is given to different extents:
Fig. 6. The Sombrero image. (a) The 3D ground truth. (b) The input image. (c), (g), (k), (o) The reconstructed depths by our method. (d), (h), (l), (p) Horn’s method. (e), (i), (m), (q) Zheng and Chellapa’s method. (f), (j), (n), (r) The Wavelet method under free boundary conditions, under boundary normals conditions, under boundary depth conditions, and under boundary normals + depth conditions (appear in this order).
1) no knowledge (free boundary);
2) surface normals are known at the image boundary;
3) surface depths are known at the image boundary;
4) both depths and normals are known at the image boundary.

As has been mentioned in Section 3, for both Horn’s and Zheng and Chellapa’s methods, natural boundary conditions should be imposed for components which are not given at the image boundary (for either depths or normals, or both). Figs. 6c-r show, respectively, the recovered depths of the four methods the four situations. The depth errors and surface normals (average error, standard deviation and maximum error) in each case have been listed in Table 2. Since the depths under free boundary and boundary normals can only be determined up to a constant shift, the average depth error here has no meaning, and the maximum error is computed as the maximum error-difference to the average error. From both the figures and the table, we can make the following observations:

1) The use of a priori knowledge has improved the accuracy of both surface depth and surface normals;
2) Depth knowledge is more powerful than normal knowledge in the reduction reconstruction errors;
3) It makes little extra improvement by using both depths and normals at the same positions;
4) In all the cases, our method gives the most accurate results, while Horn’s and Zheng and Chellapa’s methods possess almost the same accuracy.

The wavelet method has problems in producing consistent results; for instance, the use of both boundary depth and boundary normals yields even worse results than that using the depth constraint alone.

In applying a priori knowledge, it is worth mentioning that in variational approaches, like those of Horn, Zheng, and Chellapa, and the Wavelet method, the a priori knowledge serves as hard constraints, that is, the given values of the surface depths or surface gradients are set fixed at the corresponding image positions. When the a priori knowledge is not accurate, the error tends to propagate to nearby image positions. This is not the case in our parametric approach. Our method tends to find a solution which is most compatible with the intensity information and the prior knowledge. For instance, we ever set the depths at the vertical right image boundary to have one pixel position error, that is, we use the values one pixel left to the boundary as the true values at the image boundary. Our method can still
find a solution which is almost the same as that obtained with the true boundary values.

To illustrate the ability of our method to incorporate qualitative knowledge, Fig. 7a shows a real image of the Agrippa statute with the illuminant direction coming from the North-East (Agrippa right). Fig. 7b shows the reconstructed image by our method without any boundary conditions. It was found that the left cheek is lower than the right one due to self-shadows. We then use the a priori knowledge that the left and right cheeks are at the equal height by imposing the constraint \( z_l = z_m \), where \( l \) and \( m \) represents two points at the left and right cheeks, respectively. The recovered 3D shape under this constraint is shown in Fig. 7c. It can be seen that besides the achieved equal height of the left and right cheeks, the nose becomes more steep, and the shadowed part near the left corner of the mouth is improved, too. As for the use of inequality constraints, we give one illustrative example. (More examples need to be found to verify the effectiveness of inequality constraints.) For comparison reasons, we reproduce the input image and the result of the David statue in Figs. 8a and b. It can be found that the reconstruction also contains errors: the orientation of the pixels near the middle of the left boundary is wrong, due to lack of boundary conditions. Then we use the constraint that the normals at the lower two-thirds of the left boundary are toward the West, i.e., \( n_{ij} < 0 \), where \( j \) indexes the points with \((x, y)\) coordinates as \((0, y), y = 20, 21, ..., 63\). The reconstructed surface after imposing this constraint is depicted in Fig. 8c. We can see subjectively that not only the errors at the boundary points are reduced (if not completely removed), but also the part below the eye as well as the slop of eye is improved too. It should be noted that using prior knowledge in SFS is a rather interactive process; one has to find the appropriate constraints manually.

By comparing the results of our method (shown both in this and the last subsections) with those of the other methods implemented in [39], subjective judgments still indicate that our method is the most accurate one. Computationally, our method is about three to five times more expensive than the variational approaches. This is because, in variational approaches, a pixel interacts only with its four-connected neighbors; while in our approach, many Gaussians (of different sizes) participate in the minimization of the intensity error of a given pixel; that is, a pixel interacts, indirectly, with both its local neighbors (mainly through small-sized Gaussians) and its distant neighbors (through large-sized Gaussians). Although the computation at each pixel is greatly intensified, the number of iterations required for convergence is dramatically reduced, typically to several hundreds; while the variational approaches usually need tens of thousands of iterations.

6 Conclusions

In this paper, we have presented a new method of shape from shading by parameterizing the surface depth in terms of radial basis functions. The radial basis functions, controlled by their centers, widths, and weights, are deformed to minimize the intensity error and errors of prior constraints. A simple stochastic gradient method, together with an incremental scheme and multiresolution computation, has been found sufficient for effective parameter estimation. The prior constraints used in our method can be either known depth values or known surface normals anywhere in the image; they can also be qualitative knowledge about the object surface. The new computational scheme facilitates the integration of shape from shading with stereo vision: Stereo vision may provide reliable sparse depths, while shading may be used for interpolation between the depths. Experiments have demonstrated that the reconstruction accuracy of shape from shading can be significantly improved with the use of a prior knowledge. Comparisons of our method with existing ones indicate that our method can recover a more accurate surface both in the case of free boundary and in the presence of prior knowledge. The major weakness of our method seems to the more computational cost, although the parallelism of the radial-basis-function network may allow parallel implementation.

As a future research direction, the problems of interreflection [23], the estimation of source directions (e.g., by integration with SFS [38]), and especially, the modeling of natural light sources, should be dealt with since they play a very important role in making shape from shading methods applicable to natural scenes.

Acknowledgments

The authors are very thankful to Dr. Q. Zheng at the University of Maryland, Mr. P.S. Tsai at the University of Central Florida, and Dr. K.M Lee and Prof. Kuo at the University of Southern California for providing test images. Thanks to Dr. Zheng and Dr. Hiesh, the latter at the Academia Sinica, Taiwan, for kindly providing their original Fortran- or C-codes.

References


Guer-Qing Wei (S’88-M’90) received the BS, MS, and PhD degrees in electrical engineering in 1983, 1986, and 1989, respectively, from Southeast University (formerly the Nanjing Institute of Technology), People’s Republic of China. From 1989 to 1991, he was an assistant professor at the National Lab of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences. Since 1991, he has been a senior research scientist at the Institute of Robotics and System Dynamics, German Aerospace Research Establishment, Oberpfaffenhofen.

He has published papers on sensor calibration, shape from shading, and visual servoing. He holds one patent in laparoscopic instrument tracking (with Dr. Arbter). His areas of interest include computer vision and neural networks.

Dr. Wei received the Pattern Recognition Prize (a Best Paper Award) from the German Pattern Recognition Association (DAGM) and the Austrian Pattern Recognition Association (OAGM) in 1994. He was listed in Marquis’s *Who’s Who in the World*, and is a member of the IEEE Computer Society.

Gerd Hirzinger received his Dipl.-Ing degree and the doctoral degree from the Technical University of Munich, in 1989 and 1974, respectively. In 1969 he joined DLR (the German Aerospace Research Establishment) where he first worked on fast digital control systems. In 1976 he became head of the automation and robotics laboratory of DLR and got several awards for innovative technology transfer from robotics research to applications. In 1991 he got a professorship from the Technical University of Munich. Since 1992 he has been director of DLR’s Institute for Robotics and System Dynamics.

He has published about 100 papers on robotics, mainly on robot sensing, sensory feedback, mechatronics, man-machine interfaces, telerobotics, and space robotics. He was prime investigator of the space robot technology experiment ROTEX, the first real robot in space, which flew aboard the space shuttle *Columbia* in April 1993.

He is a senior member of the IEEE, vice-program and chairman of the IEEE Conference on Robotics and Automation 1994 and 1995, and program chairman of the Intelligent Robot Systems Conference in 1994. In a large number of other international robot conferences he was program committee member or invited plenary speaker. For many years he has been chairman of the German council on robot control and administrative committee member of the IEEE Society on Robotics and Automation.

In 1994 he received the Joseph-Engelberger-Award for his achievements in the robotic sciences and the Leibniz-Award, the highest scientific award in Germany.


