Summaries of Selected Shape from Shading Papers

Nathan Funk
University of Alberta
Assignment 4 for CMPUT 605

November 18, 2004

1 Zhang et al. (1999)

This paper [5] implements and compares six well-known shape from shading (SFS) techniques. They analyse the performance on a set of synthetic and real images.

The common goal for SFS methods is to recover the depth of the scene $Z(x, y)$ at all pixel locations (x, y). Furthermore the surface normal (n_x, n_y, n_z), the surface gradient (p, q) or the surface slant ϕ and tilt θ can be determined at each pixel. The gradient (p, q) is consists of the partial derivatives of the depth $(\frac{\partial z}{\partial x}, \frac{\partial z}{\partial y})$. The slant and tilt angles are related to the normal orientation with $(n_x, n_y, n_z) = (\sin\phi\cos\theta, \sin\phi\sin\theta, \cos\phi)$ for a normalized vector.

The SFS techniques are divided into four groups: Minimization, Propagation, Local, and Linear methods. At least two techniques are discussed for each of the groups.

- **Minimization:** These approaches share the feature of minimizing an energy function. The earliest techniques include Ikeuchi & Horn (1981), Brooks & Horn (1985), and Frankot & Chellapa (1988). Seven other techniques were also reviewed in this category. Since there are two unknowns for the surface gradient at each pixel, and only one value (the intensity) is known, the problem is underdetermined. Constraints need to be introduced, which are expressed as energy functions. The most common constraint is the brightness constraint which requires that the reconstructed pixel intensity matches the true pixel intensity. Other constraints include the smoothness constraint, integrability constraint, and intensity gradient constraint. Constraints can be used in combination in attempt to achieve the best results.

- **Propagation:** The first SFS technique, which was proposed by Horn, was essentially propagation technique. The main idea behind a propagation technique is that it starts estimating geometry from a location in the image, then the shape information is propagated to areas around the initial location. Horn uses ”characteristic strips” which, starting from a singular point, are expanded over the image.

- **Local:** These approaches assume that the shape is locally spherical at any point. The two primary papers in this group are by Pentland (1984) and Lee & Rosenfeld (1985). They use first and second derivatives of the intensity as input.

- **Linear:** The key characteristic for these approaches is that they linearize the reflectance map and then solve for the geometry. This is based on the assumption that the lower order components of the reflectance map dominate. The two approaches discussed in the paper are Pentland’s technique from 1988 and Tsai & Shah’s algorithm from 1992.

Tests were performed on a set of five synthetic images under three different lighting conditions. The paper discusses the results of the synthetic vase and Mozart surfaces under lighting from distant point sources (0, 0, 1) and (1, 0, 1). Tests on the real images Lenna, Pepper and Vase are also performed. The location of the light source was either known or estimated for these cases. Six algorithms were implemented and parameters were chosen to get the best results for each image individually.

The performance evaluation compared the estimated geometry to the true surface geometry (for synthetic images). For the real images, the results were evaluated based on their appearance. Each algorithm performed a little better on certain images than others, with none of them producing consistently better results than others. Overall the results were poor for synthetic images, and even worse for real images. Furthermore the results from the synthetic data did not predict performance on real data.

The authors believe that the main reasons for the lacking performance are too simplistic reflectance models, simplistic lighting models (single distant point source) and other reasons. Very few real objects are actually Lambertian, and single distant point sources are also rare for indoor scenes. They recommend that future research is focused on allowing more complex models of reflectance and lighting. Also, by including multiple cues such as stereo and range data, the results are expected to improve greatly.

2 Horn - Robot Vision chapter 11.1

This section of Horn's book "Robot Vision" [3] describes the general problem of shape from shading and reviews the traditional characteristic strips method he proposed in his earlier SFS papers. Horn introduces the problem by describing the inherent limitations of recovering shape from a single image. If the goal is to recover two parameters $p = \frac{\partial z}{\partial x}$ and $q = \frac{\partial z}{\partial y}$ for each pixel, then the single constraint from the intensity of the pixel is insufficient to determine both parameters. So it becomes necessary to introduce at least one additional constraints such as smoothness. He also notes that people must also introduce assumptions implicitly since we perceive most images as unambiguous.

The reflectance map is used to derive equations for shape recovery. It is essentially the relationship between the surface orientation described in terms of (p, q) and brightness $E(x, y)$. This reduces the irradiance equation to $E(x, y) = R(p, q)$ where $R(p, q)$ is the reflectance map. For a Lambertian surface, the radiance map is

$$R(p, q) = \frac{1 + p_s p + q_s q}{\sqrt{1 + p^2 + q^2}\sqrt{1 + p_s^2 + q_s^2}}, \tag{1}$$

where (p_s, q_s) describes the position of the light source. Instead of using only this reflectance map, Horn instead describes how to use various types of reflectance maps.

One special case are *linear reflectance maps*. Assuming that

$$R(p, q) = f(ap + bq) \tag{2}$$

where a and b are constants, and f is a strictly monotonic function with an inverse f^{-1} then

$$ap + bq = f^{-1}(E(x, y)). \tag{3}$$

This equation constrains the possible values of p and q, although it does not allow them to be determined from the intensity alone. But for a particular direction $\theta_0 = \arctan(b/a)$, Horn shows that the slope in that direction can be determined as

$$m(\theta_0) = \frac{1}{\sqrt{a^2 + b^2}} f^{-1}(E(x, y)). \tag{4}$$

So, the equation allows determining relative depth by integrating along a line in direction θ_0. To determine absolute depth, it would be necessary to know the depth at one point along the line. The profile which is established by this procedure is referred to as a *characteristic curve*.

The same derivation can be performed on rotationally symmetric reflectance maps for which $R(p,q) = f(p^2+q^2)$. This is the case when a single light source is located at the view point, or if the view point is at the top of a hemispherical sky for example. For the general case, where $R(p,q)$ is an arbitrary function the solution of the SFS problem is described by five ordinary differential equations. The solutions to these equations are paths on the surface with orientation and are called *characteristic strips*. By starting the algorithm from points with known (p,q) values, referred to as *singular points*, the solution to the ODEs can be approximated by stepping away from the singular points while propagating the known information about the surface.

Horn notes that the main deficiencies of the characteristic strip approaches is that they are sensitive to noise, and are sequential algorithms. To allow a parallel implementation, SFS needs to be approached differently. This is the main advantage the minimization approaches have over the traditional characteristic strip methods.

3 Ragheb & Hancock - SFS on Specular Objects

The early SFS methods mostly assume Lambertian surface reflection. This paper [4] attempts to also recover shading from objects featuring specular reflection. This is done using a Bayesian framework which estimates the proportion of the specular and Lambertian reflectance contribution to the intensity at each pixel in the image. SFS models that are based on a Lambertian assumption cause considerable errors in the reconstruction when they encounter specular highlights. If the highlights can be correctly identified and handled, they provide additional accurate information about the surface orientation. Unlike with Lambertian reflectance, the specular contribution allows unique identification of the surface normal orientation when the camera and light positions are known. The normal at the specular highlight is the bisector of the viewing direction and the lighting direction. For Lambertian reflectance, the intensity only constrains the normal to a cone.

The model developed in this paper estimates the angular distribution of the surface normal at each pixel. A separate probability distribution is obtained for the Lambertian and specular components. For Lambertian reflectance, the deviation from the cone is assumed to be Gaussian. The specular distribution deviating from the specular spike is assumed to follow a Beckman distribution. They adopt a maximum a posteriori probability estimation framework which is based on the work of Besag [2]. With the estimated normals, the specular reflection component at each pixel is reconstructed using the Torrance and Sparrow model. The specular component image is then subtracted from the original image, and the SFS technique is applied to the remaining diffuse reflection component.

The main assumptions about the scene are that the surface has a constant albedo and is illuminated with a single collimated light-source (equivalent to a distant point light source). According to Lambert's law the intensity at a pixel (i,j) for a location facing the light source is $E(i,j) = \vec{N}_L(i,j) \cdot \vec{L}$, where $\vec{N}_L(i,j)$ is the surface normal for Lambertian reflection and $\vec{L}$ is the light source direction. Given the pixel intensity and light source direction, the orientation of the normal is confined to a cone. To initialize the direction of the normals, the normals are aligned with the local (Canny) image gradient. Then the surface normals are subjected to smoothing to satisfy curvature consistency constraints and rotated back onto the nearest location on the irradiance cone. The direction of the normal for specular reflection $\vec{N}_S$ is simply determined at every pixel, and does not change during the iteration process. It is calculated with $\vec{N}_S = \frac{\vec{L}+\vec{V}}{\|\vec{L}+\vec{V}\|}$.

The estimation process is iterative and is broken down into five steps. The Lambertian and specular normals are initialized as described earlier. In the first step, the current normals are used to calculated a posterior mean surface normal $\vec{M}^{(n)}(i,j)$. The second step smooths the mean surface normals according to Worthington and Hancocks method. Then, the Lambertian surface normal direction is projected back

onto the irradiance cone. In the forth step, the conditional measurement densities are calculated. Given a pixel under the Lambertian model ($\omega_{i,j}^{(n)} = L$) the distribution is calculated as $p(\vec{M}_R^{(n)}(i,j)|\omega_{i,j}^{(n)} = L)$. Similarly for specular reflection $p(\vec{M}_R^{(n)}(i,j)|\omega_{i,j}^{(n)} = S)$ is determined. The smoothness priors for specular and Lambertian reflection as well as the probabilities for each mode are then calculated. In the final step, the posterior mean surface normal is updated with

$$\vec{M}^{(n+1)}(i,j) = \vec{N}_S^{(n)}(i,j)P(\omega_{i,j}^{(n)} = S) + \vec{N}_L^{(n)}(i,j)P(\omega_{i,j}^{(n)} = L), \tag{5}$$

which is nothing but the weighted average of the two surface normals maintained for the pixel.

Experiments are performed on real images of shiny objects. The results show that the technique is effective on specular surfaces and correctly separates the specular highlights from the diffuse reflections. Only slight specular highlights remain in some of the test images.

4 Ben-Ezra & Nayar - SFS on Transparent Objects

Early in this paper [1], the authors note that recovery of transparent objects is particularly difficult since many of the fundamental assumptions of computer vision are violated when transparent objects are present in the scene. Rays are not only reflected, but also refracted and can interact with multiple surfaces. Transparent objects introduce additional ambiguities to the interpretation of images, especially when only a single image is available. For this reason, additional information is used in this approach compared to most shape from shading methods which attempt to recover shape from only a single image.

The paper focuses on the recovery of a single transparent object observed from a moving camera. Assuming that the camera movement is known, and the background environment is distant, a parametric model of the object is estimated. For example for a convex lens, the focal length of the lens can be determined from a sequence of images. Recovery of arbitrary geometries is not attempted.

Refractive behaviour is related to the so called *refractive index* of materials. For a ray of light passing from a material with refractive index μ_1 to a material with an index μ_2 part of the ray is reflected and the other part reflected. Depending on the angles involved, the ray might also be completely reflected. Snell's law describes the relationship of the refractive indices to the angle between the ray and the surface normals. With θ_1 as the angle between the incoming ray and the outward surface normal, and θ_3 as the angle between the outgoing ray and the negative surface normal, Snell's law is expressed as:

$$\mu_1 \sin\theta_1 = \mu_2 \sin\theta_3. \tag{6}$$

The rays are not all necessarily contained in the same plane when passing through an object, and the possibility of the ray being completely reflected inside the object exists.

The concept of virtual features is introduced. Features that are visible from multiple views can appear to be located at virtual locations when viewed through a transparent object or reflected off a highly specular surface. For example, a feature can appear to be located behind a mirror surface although it is actually in front of the mirror.

The object's pose is also estimated through a rotation $\mathbf{R}$ and translation $\mathbf{T}$ in addition to the parameters of the object ξ. Using the assumption that the environment is distant, it is observed that all rays from a specific feature in the environment are parallel independently of where they intersect with the object. An objective function is proposed that measures the goodness of recovery given a set of tracked features. It is written as

$$f(\xi, \mathbf{R}, \mathbf{T}) = \sum_{i=1}^{k} var(o_i(\xi, \mathbf{R}, \mathbf{T})), \tag{7}$$

where o_i is the set of directions of the reversed rays associated with a feature F_i. So the variance in the ray directions acts as a penalty in the recovery. When there is no variance in the ray directions to a specific feature, then the path of the ray is assumed to be correctly reconstructed. The problem is now a multi-parameter non-linear minimization problem. The authors state that with a reasonable initial guess it can be efficiently solved.

Experiments are conducted on simulated objects as well as on real objects. For the cases published in the paper, recovery is accurate. However the choice of parametric models and initial parameter values appear to be very close to the true values. This might indicate that the algorithm takes a long time to converge or possible does not converge at all when bad initial guesses are made, as is the case with many optimization based approaches.

5 Implementation

The implementation in this assignment focuses on a relaxation method described by Horn in section 11.7 in [3]. The technique minimizes a smoothness energy function e_s and e_i which is the energy function for the image irradiance. The two functions are combined with $e_s + \lambda e_i$, where λ is a parameter that weights the errors. With use of variational calculus, a solution is expressed in terms of differential equations. A solution can also be obtained by minimizing the discretised energy functions. This approach was chosen for the assignment since it is written out explicitly in the book.

Expressions for the Lambertian reflectance function R_s and it's partial derivatives $\frac{\partial R_s}{\partial f}$ and $\frac{\partial R_s}{\partial g}$ needed to be expressed in terms of the stereographic parameters f and g. They were derived by substituting the stereographic parameter equations into the original Lambertian reflectance function $R(p, q)$ from section 10.9. These expressions were used in the iterative solution:

$$f_{kl}^{n+1} = \bar{f}_{kl}^{n} + \lambda(E_{kl} - R_s(f_{kl}^{n}, g_{kl}^{n}))\frac{\partial R_s}{\partial f}, \tag{8}$$

$$g_{kl}^{n+1} = \bar{g}_{kl}^{n} + \lambda(E_{kl} - R_s(f_{kl}^{n}, g_{kl}^{n}))\frac{\partial R_s}{\partial g}, \tag{9}$$

where $\bar{f}_{kl}^{n}$ and $\bar{g}_{kl}^{n}$ are local averages from the previous iteration, and E_{kl} is the image intensity at the current location.

Experiments were performed on an image of a sphere lit by a single distant point light source. After implementing the algorithm according the description in section 11.7.2, the results were not satisfactory. Figure 1 shows how the reconstructed surface orientations are not appropriate although they do meet the specified constraints.

To obtain appropriate results, it was concluded that an additional constraint is required. Since the normals at the occluding boundary are known, the surface orientations can be set and forced at each iteration step. The smoothness requirement then propagates the correct orientations towards the center of the sphere with each iteration. The same effect is expected to be achieved by replacing the smoothness constraint with a integrability constraint. The result of the constrained method is shown in Figure 2.

Results for light directions other than $(p_s, q_s) = (0, 0)$ did not provide accurate results. Although the general shape of sphere was recovered, symmetry was not achieved, and orientation errors are present as shown in Figure 3. This is likely due to the fact that self-shadowing is not considered in this implementation.

In the experiments, the effects of changing λ were studied. It was noted that increasing λ generally increased the rate of convergence, however could lead to divergence if λ was too high. A value of $\lambda = 0.5$ was chosen for the other experiments.

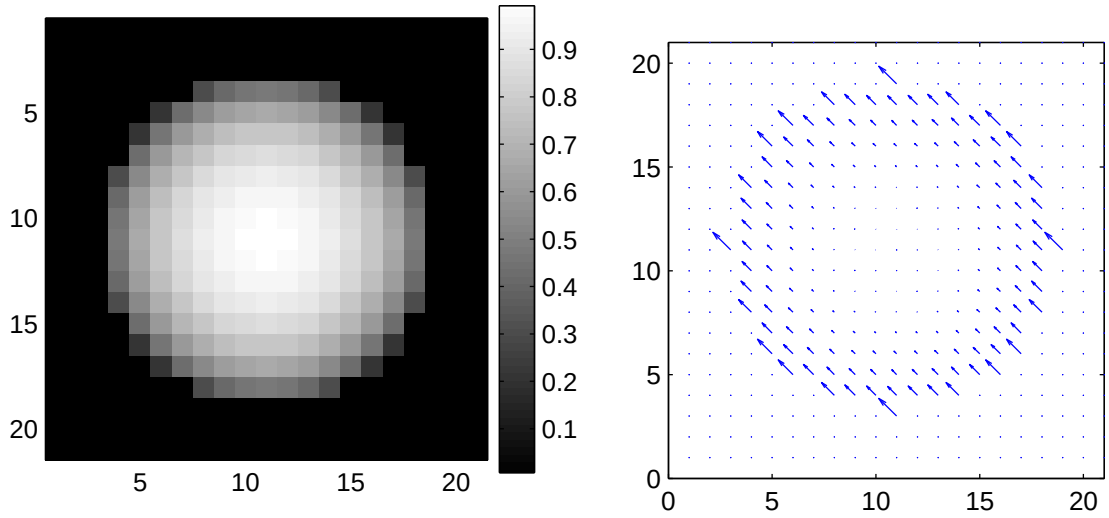

Figure 1: The initial implementation without constraining the surface orientation along the occluding boundary resulted in a false reconstruction. The orientations recovered meet all the constraints (smoothness and intensity) but can not be used to construct a surface.

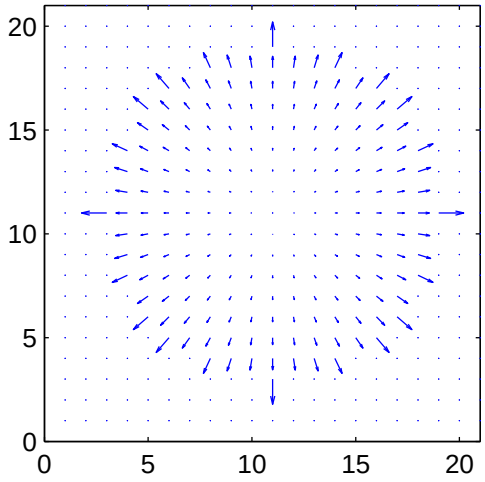

Figure 2: By adding the orientation constraint along the occluding boundary, a correct surface orientation map was obtained.

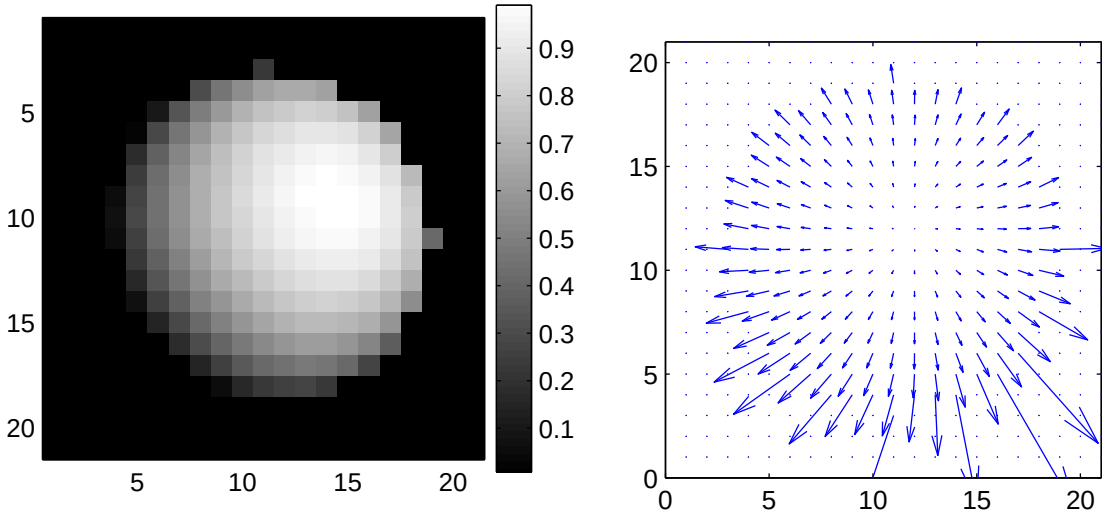

Figure 3: With an off-center light source the accuracy of the results was diminished.

References

[1] M. Ben-Ezra and S. K. Nayar. What does motion reveal about transparency? In *Proc. of the 9th IEEE Int. Conf. on Computer Vision*, page 1025. IEEE Computer Society, 2003.

[2] J. Besag. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society, series B*, 48:259–302, 1986.

[3] Berthold Klaus Paul Horn. *Robot vision*. MIT Press, 1986.

[4] H. Ragheb and E. R. Hancock. A probabilistic framework for specular shape-from-shading. In *International Conference on Pattern Recognition*, volume 3, pages 513– 516, 2002.

[5] R. Zhang, P.-S. Tsai, J. E. Cryer, and M. Shah. Shape from shading: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(8):690–706, 1999.