Scott McCloskey COMP 766
Paper Critique M. Langer and S. Zucker
"Shape from Shading on a Cloudy Day"

1 The Big Picture

In this paper, the authors outline a method for extracting the depth of an surface based on a lighting model that assumes diffuse illumination and reflections. They formalize the model in the continuous and discrete cases, and then propose an algorithm to recover depth from an image whose lighting conforms to this model.

2 The Gory Details

At the outset, the authors distinguish this paper from others in the category of shape from shading. Instead of assuming that the scene is illuminated by a point source like the sun on a clear day, the authors assume diffuse lighting. Maybe this says something about the city in which the paper was written, but this assumption corresponds to the world on a heavily clouded day. They also consider the light contributed from other points on the surface.

Formally, each point x on the surface is illuminated from a hemisphere of directions defined by the tangent plane. Some of those directions get light from the diffuse light source, and others point to other parts of the surface. Thus, the hemisphere H is partitioned into sets $\mathcal{V}(x)$ and $\mathcal{V}_c(x)$, respectively.

With the hemisphere partitioned in this manner, we can define the surface aperture, which will become important later. The surface aperture $A(x)$ is simply the percentage of directions from x that see the diffuse illuminant (the sky).

If we further assume that the cloudy sky has uniform intensity $\mathcal{I}_\mathcal{D}$, we can specify a point's output intensity as the sum of two integrals over $\mathcal{V}(x)$ and $\mathcal{V}_c(x)$.

The paper then presents a series of inequalities, most importantly the Aperture-Luminance Inequality, which specifies a lower and upper bound for the surface aperture. The authors refer to studies that indicate that the upper and lower bounds are rarely attained, and make the biggest assumption of them all. Namely, the surface aperture is taken to be the average of the upper and lower bounds. With this in place, $A(x)$ is defined as a function of $\mathcal{I}_\mathcal{D}$, $I_{out}(x)$, and ρ.

After a bit more, we get to the algorithm whose input is an image - $I_{out}(x)$ for all x. We make an assumption about ρ, but we can determine $\mathcal{I}_\mathcal{D}$ from the maximum intensity value in the image. Based on these, we can estimate the surface aperture at each x. These estimates are used to constrain possible surface arrangements.

Another constraint is necessary to recover depth from the estimated surface apertures. The Quantized Local Visibility Constraint allows us to say something about points at a lower level based on what we know about their neighbors at higher levels. Namely, if x (the lower point) can see the the diffuse illuminant in direction L, then x', a nearby point along L from x, can also see the illuminant in that direction, and x' is not on the surface. Not that x may or may not be on the surface.

The algorithm is iterative, finding the depth of the highest point, and proceeding downward in fixed distance increments. The visibility field $\mathcal{V}(x)$, and thus the surface aperture, is calculated using the Local Visibility Constraints. If this aperture is higher then the value estimated from the image intensity, then we assume the surface at point x is below the current depth. We continue in this way until the depths of each point have been set.

3 The Scott Opinion

The paper starts with one critical assumption, and piles them on as it proceeds. Here's a list of the assumptions necessary for this algorithm to work, and some comments about them:

1. The scene is illuminated by a diffuse source.
2. Surfaces are Lambertian; they have no specular highlights.
3. The diffuse illuminant (the cloudy sky) has constant intensity.
4. The surfaces are smooth, so that a normal is well defined at any point.
5. There are no atmospheric effects in free space.
6. Surfaces have constant ρ.
7. The image of the surface originates from an orthographic projection.
8. The surface has constant depth outside of the image.
9. $A(x)$ is well approximated by the average of its upper and lower bounds.

Assumption 3 is certainly not true in most images of the real world, but when is it close enough? If the clouds were thin and the sun were partially visible in the sky, we'd misinterpret shadows on flat surfaces as depressions. How much larger can the illuminant from the sun direction be without such errors occuring?

Assumption 4, while necessary for the model, does not seem obviously necessary for the algorithm. Is the normal necessary to compute $\mathcal{V}_n^*(x, y)$? We don't have enough information about the surface to compute the normal, so the answer must be no.

Assumption 7, like the third, is clearly not true in pictures of the real world. Again, we're left to ask what's close enough? Are satellite pictures of a city close enough to orthographic for this to work? We'd hope so, but it's unstated.

Assumption 8, or something like it, is necessary in order to find $\mathcal{V}_n^*(x, y)$. The exact assumption (which I haven't yet stated) can take several forms, though. The authors have chosen that the depth outside the field of view equal the depth of the node nearest the viewer. This assumption makes things simple, but one could make different assumptions that, while requiring more effort, would still work.

Assumption 9 is certainly necessary for us to even begin tackling the problem according to this model, and is necessarily coarse. This assumption seems better for some values of ρ then others, but what else would make the approximation better or worse? In particular, if we had a color image, could aperture estimates from each of the three channels be combined in such a way as to get a more realistic value? This, unfortunately, would require us to make some sort of assumption about the color of the diffuse illumination, and perhaps estimate different ρ values for each channel.

The style of the paper is generally agreeable, as things are reasonably well explained. There are a few errors, such as references to equations with the wrong number. More annoying, however, is that the table in section 8.1 and the images in figure 6 seem to follow different (and unspecified) orderings of the columns. This made it much more difficult to understand what happens when ρ isn't properly estimated.