Lecture 11: Shape-from-Shading

Adam Hawley

December 12, 2018

1 Human Shape-from-Shading

Humans can perform shape-from-shading well. They make assumptions to help solve the problem.

Shape-from-shading is also a "top down" process. High level information about the scene is incorporated into the shape estimation process. Images are interpreted to be consistent with our prior knowledge of shapes.

2 The Geometry of Surface Reflectance

Remember: Surface reflectance is described in terms of 3 unit vectors, describing the direction of the surface normal, viewer and light source. This means that intensity (I) is a function of the three unit vectors:

$$I = f(\mathbf{n,s,v,P})$$

P here refers to the material properties of the surface.

3 Shape-from-Shading

Shape-from-shading aims to estimate a surface normal for each pixel in an image, a depth map can subsequently be estimated from the normals.

4 Finding Depth From Illumination

4.1 Unknown Illumination: The Bas Relief Ambiguity

With different models, the same image can be created by moving the angle of the light source. This means that we cannot be sure of the depth.

4.2 The Lambertian Reflection Model i

Remember: the Lambertian model describes perfectly diffuse reflectance. It assumes that light is scattered equally in all directions. Intensity is given by cosine of angle of incidence:

$$I = L_d k_d \cos \theta_i$$
$$I = L_d k_d (\mathbf{n} \cdot \mathbf{s})$$

k_d is the **diffuse coefficient** where $0 < k_d \leq 1$ representing the proportion of light that is diffusely reflected by the surface (the rest is absorbed).

In a greyscale image we have one intensity value for each pixel. A surface normal is a 3D vector, so it is described by 3 values (x, y and z). But its length is fixed to one. Hence, it has two degrees of freedom (the two angles of a spherical coordinate representation). This means that we are left trying to recover two unknowns from one measurement and hence the problem is ill-posed (under constrained).

To try and get around this issue we can make simplifications by assuming:

- Diffuse coefficient $(k_d) = 1$ — A perfect white diffuse reflector

- Light source direction is known and light source intensity $(L_d) = 1$

4.3 The Convex-Concave Ambiguity

Even with known light source direction, ambiguity remains.
    SEE SLIDES FOR EXAMPLE

4.4 The Lambertian Reflection Model ii

Under the previously outlined assumptions we have:

$$I = \cos \theta_i$$

So we can recover the original angle between the surface normal and the light source:

$$\theta_i = \arccos I$$

In other words, the surface normal must lie on a cone whose opening angle is determined by the image brightness. The problem is transformed to choosing a position on the cone for the normal at each pixel.

4.5 Constraints

Surface normals at adjacent pixels are not independent. They must correspond to a real surface. Surface normals which satisfy this constraint are said to be *integrable*.

Other constraints often used include the following:

- **Smoothness** — surface normal direction should vary slowly across the surface, with no discontinuities.

- **Global convexity** — prefer convex interpretation to concave.

4.6 The Worthington & Hancock Algorithm

The Worthington and Hancock alogrithm uses the cone constraint in an iterative framework.

- Initialise normals to on-cone position.

- Enforce additional constraint (e.g apply smoothing) resulting in off-cone normals.

- Rotate normals back to closest on-cone positions.

- Repeat until convergence.

4.6.1 Initialisation

Worthington and Hancock use an itialisation in which the normals line up with the local negative image gradient. This has the effect of picking the darkest point on the cone. This idea is consistent with an assumption of convexity.

4.7 Deep Learning

It should be noted that deep research into the application of deep learning is being undertaken. Neural networks have produced models which can compute depth maps in milliseconds and using fewer assumptions than we are discussing. For example they are able to estimate the lighting intensity.

4.8 Shape-from-Shading Images Conclusions

- Shape-from-shading is still an unsolved problem.

- Fails on real-world images.

- Even on synthetic images, surface can be highly distorted.

A more practical alternative is to use **photometric stereo**. This involves taking multiple images from the same viewpoint but with varying illumination. Each image provides a partial constraint on the surface normal direction. They can all then be combined.

4.9   Photometric Stereo

Two light source directions gives us two cones and two points of intersection. Two points of intersection means that there is two possible normal directions. Three light sources uniquely determines the normal. However, because of noise, this may not be an exact solution. This means that in reality it is best to use more than three light sources and not compute it exactly but use something similar to a least squares solution.

One image gives the following equation for each pixel from the Lambertian model equation:

$$I = L_d k_d (n_x s_x + n_y s_y + n_z s_z)$$

We can factor the diffuse coefficient into the normal since these are our unknowns and light source intensity into the light source vector since we know both of these:

$$I = (k_d \mathbf{n}) \cdot (L_d \mathbf{s})$$

In other words, we have a linear equation in the scaled surface normal. The three unknowns are:

$$N = [\bar{n}_x \bar{n}_y \bar{n}_z]^T$$

Hence, we need at least three images. With three images, we have 3 simultaneous equations for each pixel:

$$I_1 = \bar{s}_{1x} \bar{n}_x + \bar{s}_{1y} \bar{n}_y + \bar{s}_{1z} \bar{n}_z$$
$$I_2 = \bar{s}_{2x} \bar{n}_x + \bar{s}_{2y} \bar{n}_y + \bar{s}_{2z} \bar{n}_z$$
$$I_3 = \bar{s}_{3x} \bar{n}_x + \bar{s}_{3y} \bar{n}_y + \bar{s}_{3z} \bar{n}_z$$

Subscripts represent image number, bar means the vector has been scaled by diffuse coefficient or light source intensity. I and s quantities are known, n are unknown. One can solve for n by rearranging and combining equations solving for:

$$N = [\bar{n}_x \bar{n}_y \bar{n}_z]^T = [k_d n_x k_d n_y k_d n_z]^T$$

We form a matrix of the scaled light source vectors. Assuming we have k images, this will be of dimension k times 3:

$$S = \begin{bmatrix} s_1^T \\ \vdots \\ s_k^T \end{bmatrix}$$

4