CSE 252A: Photometric Stereo

Lecturer: David Kriegman

Scribed by Owen Jow on October 23, 2018

1 Photometric Stereo

- *Binocular stereo*: characterized by multiple viewpoints, dynamic scene, one lighting condition.
- *Photometric stereo*: characterized by one viewpoint, static scene, multiple lighting conditions.

1.1 Overview

In the process of photometric stereo, we reconstruct a 3D surface patch from multiple images of the surface under different lighting. Typically, we want a constrained or assumed setting in which the surface is diffuse, indirect illumination can be ignored, and camera response is linear in radiant exitance. Then there is a simple equation for pixel brightness as a function of the visible surface normal at a pixel. We can solve for all of the normals and then integrate them to recover the depth at each pixel.

1. We can obtain images by taking pictures of a static scene with a static camera, with different lights in various configurations of on and off. For our purposes, we will assume distant lighting.
2. Given the k images, we then (independently) estimate a surface normal for each pixel location in the scene.
3. Finally, we integrate the normals to estimate depth across the photographed surface.

1.2 On Surface Normals

A surface normal is a unit vector orthogonal to the tangent plane of the surface. Therefore, we can think of it as a point on a unit sphere. Note that in photometric stereo, we can only estimate normals that are facing the camera, i.e. we only care about normals with directions within the half of the sphere that is oriented toward the camera. Since we only care about directions over a hemisphere, we can then parameterize a normal in terms of a **slant** (p) and a **tilt** (q) ["*in gradient space*"].

- These are the partial derivatives of a depth function w.r.t. x- and y-axes.

  We're essentially trying to reconstruct a depth map, i.e. depth $f(x, y)$ as a function of x, y coordinates. In this vein, we parameterize the imaged surface $\mathbf{s}(x, y)$ as $(x, y, f(x, y))$. Then

$$\frac{\partial \mathbf{s}(x,y)}{\partial x} = \left(1, 0, \frac{\partial f}{\partial x}\right) = (1, 0, p)$$
$$\frac{\partial \mathbf{s}(x,y)}{\partial y} = \left(0, 1, \frac{\partial f}{\partial y}\right) = (0, 1, q)$$

These two vectors define a local coordinate system on the surface that span the tangent plane. *Thus the normal* $\mathbf{n}$ *is the vector orthogonal to these two vectors (their cross product).*

$$\begin{aligned}\mathbf{n} &= \frac{\partial \mathbf{s}}{\partial x} \times \frac{\partial \mathbf{s}}{\partial y} \\ &= \left(-\frac{\partial f}{\partial x}, -\frac{\partial f}{\partial y}, 1\right) = (-p, -q, 1)\end{aligned}$$

If we want a unit vector, we can normalize:

$$\hat{\mathbf{n}} = \frac{1}{\sqrt{p^2 + q^2 + 1}}(-p, -q, 1)$$

The main results: (1) p and q are the partial derivatives of depth, (2) they encode the normal.

1.3 On Prior Knowledge

The amount of information we have about the BRDF and lighting has a large impact on our photometric stereo process. We might (1) know the BRDF, which is arbitrary, (2) assume a Lambertian BRDF with known lighting directions, (3) assume a Lambertian BRDF with unknown lighting directions, or (4) have an unknown arbitrary BRDF with known lighting directions.

1.4 General Known BRDF

What is the brightness ("image irradiance E") we measure at pixel (x, y)? Assuming illumination by a distant, unit-strength light source, it's a function of the BRDF at corresponding surface point A, the surface normal at A, the brightness and direction of the light source, and the direction from A to the camera lens. Basically this is all the information we use to evaluate the BRDF.

In this case, we know everything except the normal.

To aid in our estimation process, we will use a **reflectance map** $R(p, q)$, which tells us the image irradiance under the assumptions of distant lighting, orthographic projection, and a constant BRDF. It's specific to a lighting direction and a BRDF, and can be measured by taking a photo of a sphere.[1]

- If we have a brightness at a pixel, we can look at the contour plot of the reflectance map and figure out which (p, q) values would lead to that brightness.
- Then, if we take another image with a different lighting direction, there's a different reflectance map and we can check the contour corresponding to the pixel brightness again. The true (p, q) should be at one of the intersections of these contours (there could be multiple intersections).
- Then we can add another image and reflectance map, and hopefully figure out an exact (p, q). In this way we get the normal from brightnesses across three images.

And we repeat this across the image to estimate a full normal field.

The full process is

1. Construct reflectance maps $R_1(p, q)$, $R_2(p, q)$, $R_3(p, q)$ for each light source direction.

[1] We compute R analytically from the photograph by fitting a circle to the photographed sphere, then using the sphere equation $x^2 + y^2 + z^2 = r^2$ to find a normal given a point (x, y) in the image, and finally mapping this normal to the brightness at (x, y).

2. Acquire three images $E_1(x, y)$, $E_2(x, y)$, $E_3(x, y)$ with these lighting directions.
3. For each pixel (x, y), equate $R_i(p, q)$ with $E_i(x, y)$ (three equations and two unknowns). We can solve this overdetermined system of equations for p, q i.e. the normal at pixel (x, y).
4. Continuing, we can estimate the surface normals for all pixels.

But this isn't yet depth. The final step is to go from the normal field to a 3D mesh [the surface $f(x, y)$]. Prof. Kriegman calls this "integrating the normal field."

This means solving a system of partial differential equations

$$\frac{\partial f}{\partial x} = p(x, y)$$
$$\frac{\partial f}{\partial y} = q(x, y)$$

for $f(x, y)$. We can do so in a scanline fashion by walking in each direction and setting each f value to a cumulative sum of partial derivatives in that direction (initializing the first value to 0, e.g.).

1.5 Photometric Stereo with Known Lighting & Lambertian Surfaces

Assuming a Lambertian surface, the brightness of pixel (x, y) is

$$\begin{aligned} E(x, y) &= [\rho_d(x, y)\mathbf{n}(x, y)] \cdot [s_0 \mathbf{s}] \\ &= \mathbf{b}(x, y) \cdot \mathbf{c} \end{aligned}$$

for $\rho_d(x, y)$ the albedo of the surface projecting to (x, y), $\mathbf{n}$ the surface normal, s_0 the light source intensity, and $\mathbf{s}$ the direction to the light source.

The unknown parts of the equation are the albedo and the normal, i.e. $\mathbf{b}$.

We can solve for $\mathbf{b}$ from 3+ equations (3+ images):

$$E_1 = \mathbf{b}^T \mathbf{c}_1$$
$$E_2 = \mathbf{b}^T \mathbf{c}_2$$
$$E_3 = \mathbf{b}^T \mathbf{c}_3$$

since E_i and $\mathbf{c}_i$ are known. Then the normal is $\mathbf{b}/\|b\|$ and the albedo is $\|b\|$.

(This is all for the per-pixel case.
But we can again do this for all pixels across the image to get a normal field.)

1.6 Photometric Stereo with Unknown Lighting & Lambertian Surfaces

"What is the set of images of an object under all possible lighting directions?"

An image with 10,000 pixels is a point in a 10,000D space (i.e. with a coordinate axis for each pixel). One question is, what is the space of images of some subject? Or, what is the subset of all possible images of the subject in this high-dimensional space?

Under a Lambertian model, where each image $\mathbf{x}$ (all the pixels stacked into a vector) is $\max(\mathbf{B}^*\mathbf{c}, 0)$,[2] the set of images of a surface under *all lighting conditions without pixels in shadow* is the 3D linear

[2] If $\mathbf{B}$ is the matrix with per-pixel $\mathbf{b}$ vectors as rows, then $\mathbf{B}^*$ is just $\mathbf{B}$ up to some linear transformation.

subspace $\{\mathbf{x} \mid \mathbf{x} = \mathbf{B}^*\mathbf{c},\ \forall \mathbf{c} \in \mathbb{R}^3\}$. Given three images, we can construct the basis images (in $n \times 3$ matrix form[3] $\mathbf{B}^*$) of that 3D linear subspace. Then, with the basis images, we can determine what the object will look like under any lighting condition.

To estimate the subspace $\mathbf{B}^*$ from 3+ images, we can use SVD. We stack image vectors into a data matrix $\mathbf{D}$ (e.g. a $10{,}000 \times 6$ matrix for six images), then take the SVD of $\mathbf{D}$ to get a basis of three images (the first three columns of the left matrix in the decomposition). At last we can relight the images to simulate different lighting conditions.

The set of images of an object under all lighting conditions, i.e. when you account for shadowing, lie on the boundary or in the interior of a convex cone. (Images constructed under one lighting condition lie on the outside of the cone. Images constructed under two or more lighting conditions lie on the interior.) Is this cone unique? Can multiple different objects produce the same set of images? Is there an object which looks just like another object under all lighting conditions?

The answer to this is yes.

The $\mathbf{B}$ matrix can shear, squash, or extrude the surface. And no longer how you light one version of the object, there's a way to light a different version of the object to make it look exactly the same. So we can only recover a surface up to one of these *Generalized Bas-Relief transformations* if we don't know where the lights are coming from. We can always get the same image in different ways by switching up the lighting condition or the version of the surface.

Hence if we want to solve for that surface, we need the lighting condition to be fixed. We can *almost* get shape from photometric stereo without knowing where the lights are! But an ambiguity remains.[4]

[3] n is the number of pixels in the image, e.g. 10,000 in the previous and next examples.

[4] In some conditions we can use other things to resolve this ambiguity, which is only three numbers.