EE4212 Computer Vision

Assignment 2: **Photometric Stereo**

Vo Nguyen Khang - U084040W

***Abstract*— Photometric stereo is a technique to recover local surface orientations, i.e. surface normals, from images under different illumination conditions. In this project, with each set of multiple images of an object captured under different directional illuminations and the same viewpoint, the two main steps are performed: Calibration of illumination direction, and Estimation of normal and albedo. There are three basic results for each example data: a) a normal map linearly encoded in RGB; b) an albedo map; c) a re-rendered picture of the object with your recovered normal and albedo under illumination direction that is the same as the viewing direction.**

I. PHOTOMETRIC STEREO ALGORITHM

A. Calibration

In this calibration step, the lighting intensity and direction as well as the camera response function are determined. For this project, all the given images are linearized, i.e. the pixel value is the irradiance. Therefore, the remaining work in this step is to balance the light intensities in all the test images, and to find the lighting direction in each test image. The metal sphere and the white matte Lambertian sphere have been included in all the test images for this calibration purpose.

1) Balance light intensities: First, the matte Lambertian sphere is used to find the light intensity in each image, and then the light intensities in all the test image are balanced. For a Lambertian surface, the observed intensity is determined by the equation $L_o = L_i \rho(\mathbf{N} \cdot \mathbf{L})$, which does not depend on viewing direction. Furthermore, when $\angle(\mathbf{N}, \mathbf{L}) = 0$, L_o reaches its maximum $max(L_o) = L_i$, i.e. the brightest point on the Lambertian sphere should have a normal direction coincident as the lighting direction. Hence, we can obtain the lighting intensity from the pixel value of that point.

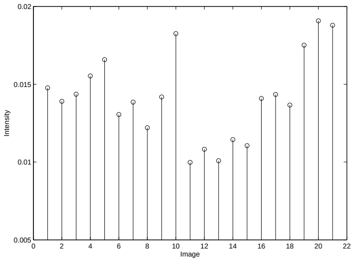


Fig. 1: Uncalibrated intensities

After having the light intensities, the values are balanced by scaling with an appropriate factor, so that they are all equal to the maximum intensity.

2) Calibration of illumination direction: The metal sphere has a bright test spot on the surface where specular reflection occurs. For specular reflection, the observed intensity is $L_o = L_i$ iff $\mathbf{V} = \mathbf{R}$. Since the highlight can only be seen at $\mathbf{V} = \mathbf{R} = [0, 0, 1]$, the lighting direction can be calculated from the equation $\mathbf{L} = 2(\mathbf{N} \cdot \mathbf{R})\mathbf{N} - \mathbf{R}$. Because the light is pointing toward the sphere, the actual equation is negated.

Assume the center of the sphere is (c_x, c_y), the sphere's radius is r, and the highlight spot in the image is (x, y), the normal of the sphere at the highlight location is $N = (x - c_x, -(y - c_y), \sqrt{r^2 - (x - c_x)^2 - (y - c_y)^2})/r$. The y coordinate is negated because the y-axis points down in image space but it is preferred to have the y-axis point up in three-dimensional space. The center of the sphere is calculated from the centroid of the mask circle, the radius is calculated from the size of the mask, and the highlight point is the center of the brightest region inside the mask.

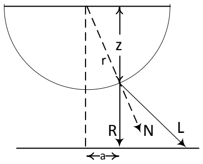


Fig. 2: Finding z

B. Estimation of normal and albedo

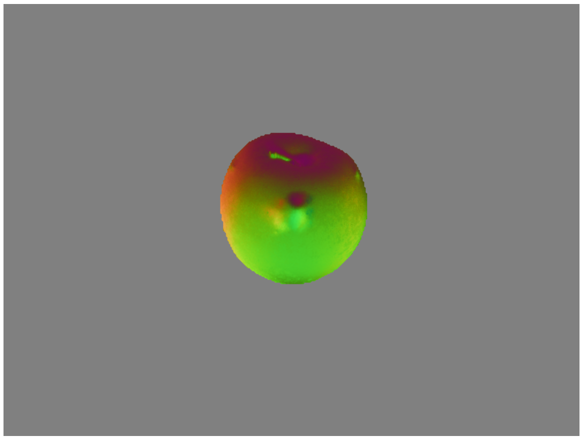
The observed pixel intensity of a Lambertian surface is $I = k_d \mathbf{N} \cdot \mathbf{L^T}$ where k_d is the surface albedo, $\mathbf{L}$ is the lighting direction, and $\mathbf{N}$ is the unit surface normal ($\mathbf{L}$ and $\mathbf{N}$ are unit vectors). To solve for $\mathbf{N}$ and k_d, at least three images (with known $\mathbf{L}$) are needed. However, in practive, more than three images are used to minimize error. Set $\mathbf{G} = k_d \mathbf{N}$, the linear least square system is formed $I_i = \mathbf{G}\mathbf{L_i^T}$. To deal with shadows,each equation is given a weight, with larger weights are given to brighter pixels. The final linear least square system needed to be solved is $I_i^2 = \mathbf{G}\mathbf{I_i}\mathbf{L_i^T}$. SVD is a powerful method to solve this type of problem. After having $\mathbf{G}$, the albedo and normal are obtained directly from the equations $k_d = ||\mathbf{G}||$, $\mathbf{N} = \mathbf{G}/k_d$. The normal map is obtained by shifting by 1 and scaling the normals by 0.5. The albedo is also scaled for visualization purpose.

C. Reconstruction depth from normals

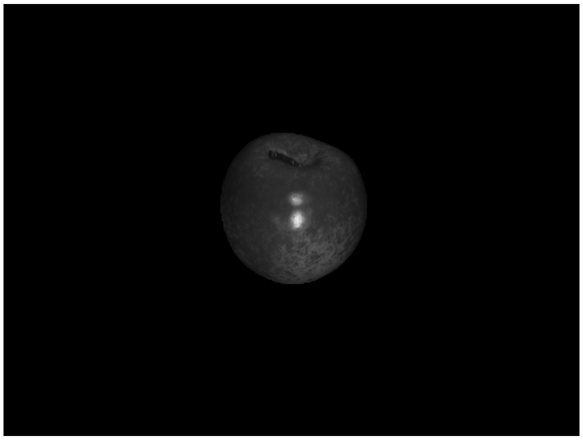
The basic idea is that If the normals are perpendicular to the surface, then they'll be perpendicular to any vector on the surface. Based on that idea, for each pixel (x, y), two equations can be formed for the top pixel $(x, y + 1)$ and for the right pixel $(x + 1, y)$: $N_z(z_{xy} - z_{x+1,y}) = N_x$ and $N_z(z_{xy} - z_{x,y+1}) = N_y$. In the end, we have to solve the linear least square system with as many unknowns as the number of pixels, for example, with a 800×600 image, there are 480,000 unknowns, and the number of equations doubles that number. One way to solve is to use sparse matrix. However, depth reconstruction is not strictly required in this assignment.

II. RESULT AND DISCUSSION

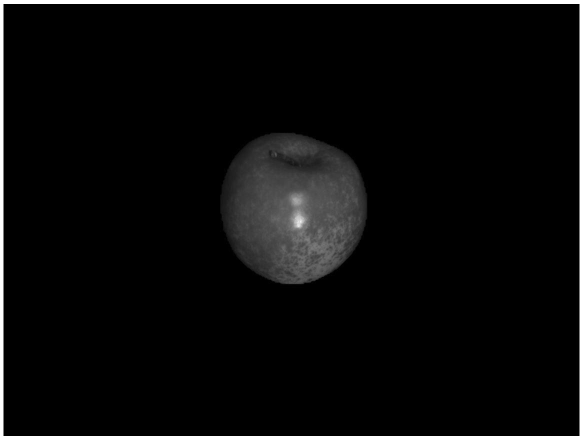
The algorithm is tested on two sets of images. In each set, there are three results including the normal map encoded in RGB, the albedo map, and the re-rendered picture of the



(a) Normal map



(b) Albedo map



(c) Re-rendered image

Fig. 3: Image Set I - Apple

(a) Normal map

(b) Albedo map

(c) Re-rendered image

Fig. 4: Image Set II - Elephant

object using the recovered normal and albedo under illumination direction that is the same as the viewing direction ($[0, 0, 1]$).

A. Limitations of photometric stereo

In this assignment, Lambertian photometric stereo is used; therefore, the most obvious limitation is that it cannot handle shiny objects, such as in Image Set I (apple). For example, in Figure 3(a), the normals at the small shiny part of the apple cannot be estimated. Besides, careful setup and calibrations on camera and light source are required. Furthermore, Lambertian photometric stereo is also not robust against shadows, inter-reflection and noise. Another limitation is the calculation time, which is rather slow, especially for large images.

B. Suggestions

Currently, all the images are converted to gray-scale images before processing. However, all the processing can be done on each separate channel of the image, and the resultant normal maps and albedo maps can be combined together to improve accuracy. The depth reconstruction step is quite interesting, but due to limited time, the code is still unfinished and cannot be included in the report.