Computer Vision 1-Assignment 1

Michael Mo
University of Amsterdam
michael.mo@uva.student.nl

Changxin Miao
University of Amsterdam
changxin.miao@uva.student.nl

1 Introduction

For this assignment, in the first part, we showed that photometric stereo can be used to estimate albedos, outward normals and surface height maps with a series of images. Multiple experiments are then conducted to understand the characteristics of the technique and its performance.
In the second part, we modeled different color spaces, investigated the theories behind those models and respective advantages. Then we tried to decompose the intrinsic image into reflectance and shading model, which could help us to analyses the initial color accurately. In the end, we implement the gray-world algorithm and build the color consistency models to automatically correct color distribution of a picture.

2 Photometric Stereo

2.1 Estimating Albedo and Surface Normal

In this part we estimate the albedo and surface normal of a 3d object, where we are only given some images of the object. The object and camera are assumed to locate in the same location, while the illumination comes from different directions in each image. To estimate the albedo and surface normal, we implement the first part of algorithm 5.1 in section 5.4 of Computer Vision: A Modern Approach.

Q1.1 Albedo is the surface reflection of the object, which is a factor indicating how much of the light is reflected for that surface. The surface of the object we look at consists of four quarters: 2 of them are white and 2 of them are black. White color is supposed to reflect more light than black, so we expect that the 2 whiter quarters have much higher albedo values than the 2 darker quarters in the albedo image. We indeed observed it from Figure 1 that the more whiter quarters have higher albedo, and the darker quarters a lower albedo.

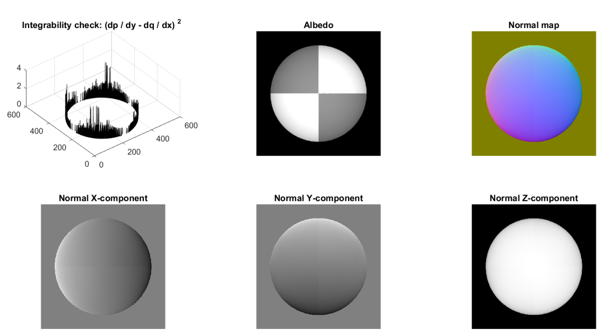

Figure 1: Test of integrability, albedo and normal for GraySphere5 with shadow trick.

Submitted to 31st Conference on Neural Information Processing Systems (NIPS 2017). Do not distribute.

Q1.2 For each point the algorithm requires us to solve the system of linear equations$Vg = i$ (or $IVg = Ii$ with shadow trick) for g. Therefore, to determine the 3 unknowns in g we will need to have at least 3 (independent) equations, which means we need at least 3 images in order to estimate the albedo and surface normals. (If we have more than 3 images, then we get more than 3 equations and have an overdetermined system. In that case, there is probably no exact solution, but then we will calculate the least squares solution.)

To answer how many images are necessarily, we note that the number of equations is not the most important, but whether you have multiple images where the light source is not positioned close to each other or not (especially when using the shadow trick). To demonstrate this: In Figure 2we have used 10 images where the light source was mostly coming from the left side of the object. This gives worse results then when we had used only 5 images.

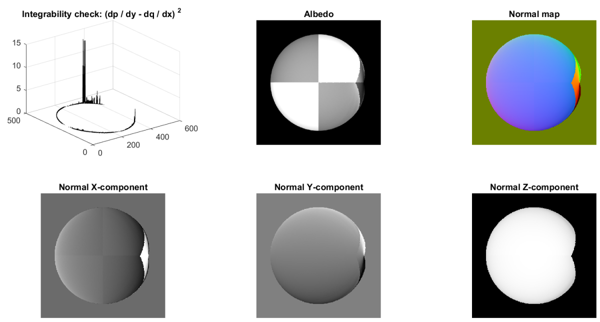

Figure 2: Test of integrability, albedo and normal for 10 out of 25 images with shadow trick.

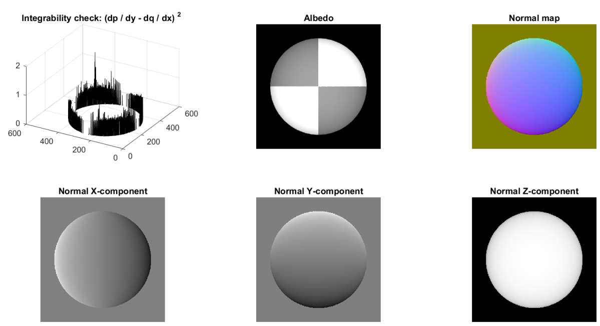

Figure 3: Test of integrability, albedo and normal for GraySphere25 with shadow trick.

Q1.3 The algorithm we worked on to estimate the albedo and surface normal is based on the lambertian model of the BRDF (bidirectional reflectance distribution function). This means we assume that for a fixed point p on the surface of the object, we have that the only source of light for p comes from the point light source (which is assumed to be distant), after which some light then gets reflected into the viewer's eyes (the camera in this case).

Now we suppose that for an image i_n, p is a point which lies in a shadow. In the model, the light does not go to that point and cannot be reflected to the camera. So for that image i_n and point p, we do not want to use any information (equation). Therefore we multiply the equation on both sides with the image value point p has in image i_n. Since that will be 0, it means that equation will be canceled out. In the case where a point does not lie in a shadow, then it does not matter since we are then multiplying both sides by just some number, which does not matter for solving the equation. Therefore instead of solving $i = Vg$, we solve $Ii = IVg$. This is the shadow trick, and it gets rid of equations derived from points in shadowy regions. If we do not use the shadow trick, we can obtain the following results in Figure 4.

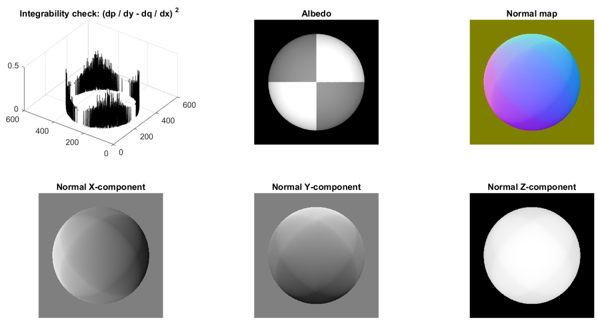

Figure 4: Test of integrability, albedo and normal for GraySphere5 without shadow trick.

The difference is explicit when we look at the normal-X and normal-Y components in the case of using only 5 images. When we use 25 images, there is almost no difference. The reason is that with 25 images, we have more than enough equations for all the points.

2.2 Test of Integrability

The end goal is to create a surface height map, which enables us to estimate the values of f. We do that by first estimating the derivatives p=df/dx and q=df/dy at each point of the image.

To evaluate the calculated estimates make sense, we can do a test of integrability, meaning we check whether $\frac{d^2f}{dxdy}$ equals $\frac{d^2f}{dydx}$. They should be the same, but could differ a small amount due to measurement errors and rounding.

Q2.1 In order to calculate these derivatives, we make use of the unit normals at every point which we calculated in part 1.1.

Q2.2 To calculate those second order derivatives, we need to find the derivative of p with respect to y, and of q with respect to x. As approximation for that derivative, we make use of its neighbors:

$$\frac{dp}{dy}(a, y) \approx 0.5 \cdot (p(a, y + 1) - p(a, y - 1)), \quad \frac{dq}{dx}(x, b) \approx 0.5 \cdot (q(x + 1, b) - p(x - 1, b))$$

We can perform a test of integrability by checking the values of $SE = (dp/dy - dq/dx)^2$ for each point (x,y). It is clear that most of them are close to zero ($<$ 0.005=threshold), but not all of them. We believes that there are lot of errors at borders of the object (or where the object is not smooth). This can be correct because at those places the function will jump a lot. Since we make use of a sampling of the function , we cannot calculate the exact partial derivative but only by means of approximation. Thus,we expect in those places there will be a higher chance and magnitude in error. Indeed in Figures 1and 3we see that in both cases (5 and 25 images) the test of integrability fails at the border of the object.

2.3 Shape by Integration

To create a surface height map f of the object we estimate f using the estimated partial derivatives (p/q) of f as follows:

$$f(a + 1, y) \approx f(a, y) + p(a, y), \quad f(x, b + 1) \approx f(x, b) + q(x, b)$$

and assume a starting point f(0,0) with value 0. This also means that to estimate f(x,y) we can take multiple paths. For example: y-coordinate first, and then x-coordinate (column-major). Or x-coordinate first, then y-coordinate (row-major). Or take multiple paths and take the average estimated value.

Q3.1 Before we create the surface height map, we perform an extra intermediate step: After the test of integrability, we know some pixels where the estimated partial derivatives are "doubtfull". To handle those cases, we replace those values by zero. This means we make the assumption that f stays constant around that pixel. (We could do interpolation of "not-doubtfull" values instead, but replacing them by zeros already gave good results.)

We use the 5 images of GraySphere5 with the shadow trick and check the results of using column-major or row-major paths. Both return satisfactory outcomes. However, at the transition between the whiter quarter and the darker quarter, we do see a tiny mishap. In general if we do column-major then (because the surface height is calculated per row) we expect the surface map to be smooth for each horizontal line but not necessarily when we look vertically. For row-major the opposite is expected. In part 1.4 we will have a more complicated object, and there we will this difference much more.

Q3.2 When we take the average of the two surface height maps, the result again looks very similar. In general, averaging the two will smoothen both of them and consequently get rid of the sudden jumps horizontally or vertically in the surface map. Using 25 images of the gray sphere we get a better result where the tiny mishap is gone. In part 1.4 we will see a much bigger difference.

2.4 Experiments with different objects

Q4 We see in Figure 5that the albedo result and the surface map generated of the monkey face looks quite good in general. There are more outliers, but that is expected since compared with the sphere, the monkey face is much less smooth and the function to be reconstructed has much more steep jumps.

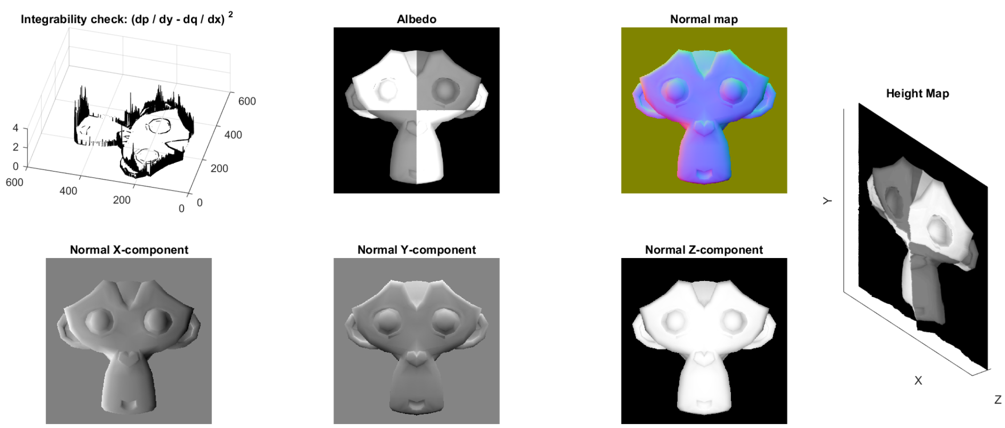

Figure 5: Albedo, normal and surface height map for monkey gray.

Q5 We now handle color images and for the albedos this is no problem. We simply interpret the albedos, as how much they reflect either red or blue or green colors, and the procedure to calculate the albedos then stays the same. (We calculate it for each channel one time so that albedo becomes RGB too).

If we simply make use of the R value, then we can get problems when we want to calculate the normals. The reason is that the R value might be zero while in the gray scale image it is not zero. This happens when a point of the surface absorbs all the red light. To overcome this problem, we can simply combine the 3 channels to a single channel by taking some linear combination of the RGB values. We choose to take the average of the three values.

The results for colorsphere and colormonkey are in Figures 6 and 7.

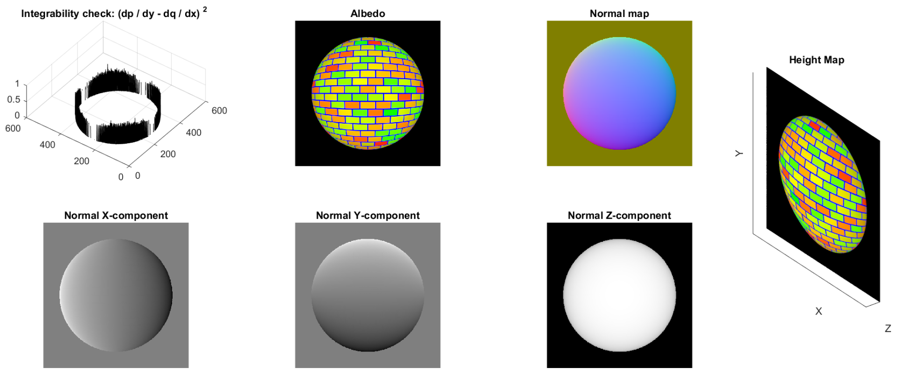

Figure 6: Albedo, normal and surface height map for sphere color.

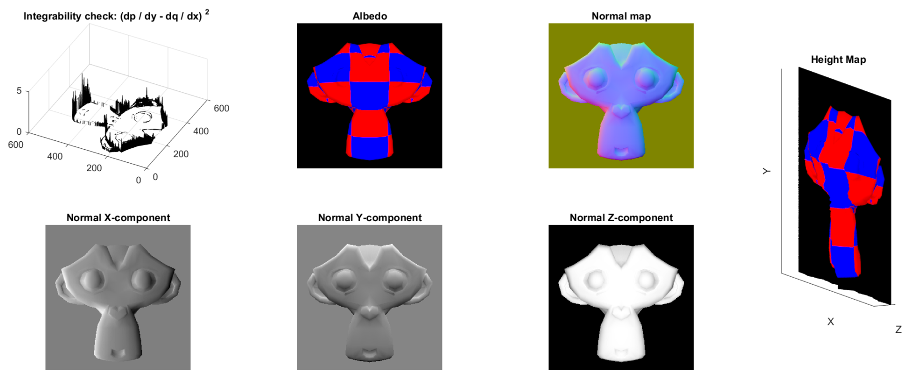

Figure 7: Albedo, normal and surface height map for monkey color.

Q6 We have 64 images and we compare the surface height map for column-major, row-major cases or take the average of the two paths . In Figure 8we see that when we do column-major, each horizontal line of the surface map looks quite smooth, but vertically it is not. The reason is already given in part 1.3. Similarly, for row-major we get that vertically it looks smooth, but horizontally not. In the average case, we have a more smoothed version in both directions.

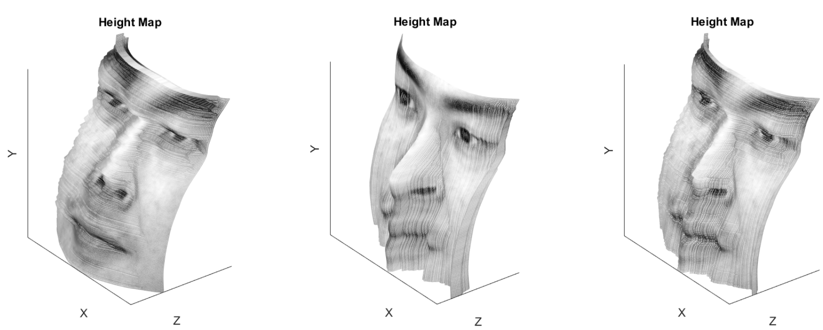

Figure 8: Column-major (left), row-major (centre) and average (right) integration paths.

In the dataset of 64 face images, there are some problematic images, which do violate the assumptions of the used method (i.e. the light is too close or is not a point source, or there is too much obstruction or shadow). We removed some of the problemsome images and did a reconstruction, but we could not see any significant difference.

3 Color Spaces

3.1 RGB Color Model

Based on human perception of colors, more specifically, the trichromacy, RGB model is used as a basic of digital cameras. The trichromacy indicates that the organism's retina contains cone cells which help us to identify different colors. The interactions among these cone cells mediated the trichromatic color vision and enables humans and some other animals to see different colors. In order to match the human eye mechanism, the digital camera and television industry chose RGB system to display colors.

A digital camera replaces film with a sensory array. Then the light activates imaging sensors array such as CCD and CMOS with a certain time period. Subsequently, the cells in arrays are light-sensitive diode that convert photons into electrons. Afterwards, the NTSC system is often used in digital cameras and videos to transform XYZ in generated by electrons into RBG. It uses the C illuminant as the reference white color. The primary RGB values are set with certain values. In order to display RGB values, color filter array were used in mosaic pattern(e.g.Bayer patterns) in digital cameras in order to display the color values.

3.2 Color Space Properties

3.2.1 Opponent color space

One of the properties for the RGB system is that the values of three colors in this system are highly correlated. If we decor-relate the RGB values, opponent color space will be generated. Opponent color space is a HVS model consists of three elements.-O1 is luminance component -O2 is the red-green channel -O3 is the blue-yellow channel. Humans have difficulty in perceiving reddish-greens and bluish-yellows. With this model, it will be rather easy for them to identify these colors. After the transformation, the yellow paprikas in the background are difficult to read while black shadow becomes more prominent. The opponent color space model is device dependent and not perceptually uniform.

Source:`chrome-extension://oemmndcbldboiebfnladdacbdfmadadm/https://engineering.purdue.edu/~bouman/ece637/notes/pdf/Opponent.pdf`

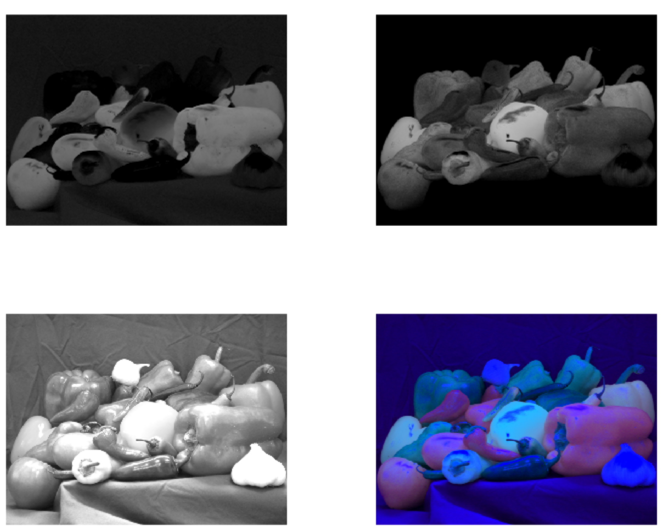

Figure 9: From RGB to the opponent and individual channel

3.3 Normalized RGB (rgb) Color Space

In the RGB system, color pixel values are highly dependent on the light source characteristics and the object geometry. If we would like to observe the RGB values independent on surface orientation, illumination, direction, and illumination density, it is quiet important for us to normalize the RGB color space, as Lights and shadows might cause distortions in an image. If we normalize the RGB values, it's possible for us to get rid of those elements. The effect could be clearly observed by the below figure. After the transformation, the white and black colors become indistinguishable. For example, the onions and garlics become dark black in the last picture.

Source:`http://aishack.in/tutorials/normalized-rgb/`

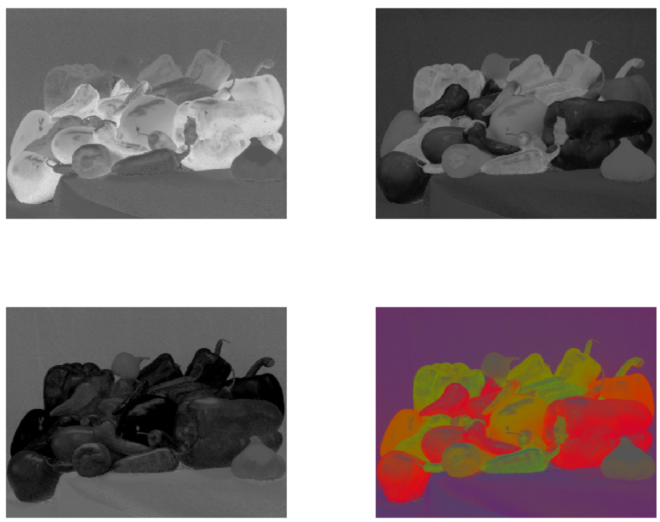

Figure 10: The transformed picture and individual channel

3.4 HSV Color Space

The HSV color space consists of three channels: the Hue, Saturation and the intensity. The Hue channel represents the dominant length of the spectral power distribution. The Saturation defines the amount of color or purity of color which could be clearly identified. Moreover, intensity is the brightness of the color. Unlike RGB, HSV separates the image intensity , from chroma or the color information. Hence HSV color space can be quiet useful if researchers want to extract certain features such as Hue, Saturation of the intensity from the current picture.

3.5 YCbCr Color Space

Human eyes have different sensitivity to color and brightness.The YCbCr stands for different meanings, Y: Luminance, Cb: Chrominance-Blue and Cr: Chrominance-Red. Luminance is the same as intensity, Cb is strong with image containing the color of blue and Cr is emphasized in places of occurrence and reddish colors. Based on the physical framework of the eye, we don't really need to retain all color information in our arrays, but part of information that eyes might be sensitive to. There are also different transformation schemes designed for different purposes. $YCbCr$ is utilized to separate a luma signal (Y) and other two chroma components (Cb and Cr). The luminance(Y) could be stored in high-resolution or be transmitted to high band-width, while (Cr) and (Cb) could be low-band reduced or compressed. In this way, the system efficiency could be enhanced significantly.

Source:`https://makarandtapaswi.wordpress.com/2009/07/20/why-the-rgb-to-ycbcr/`

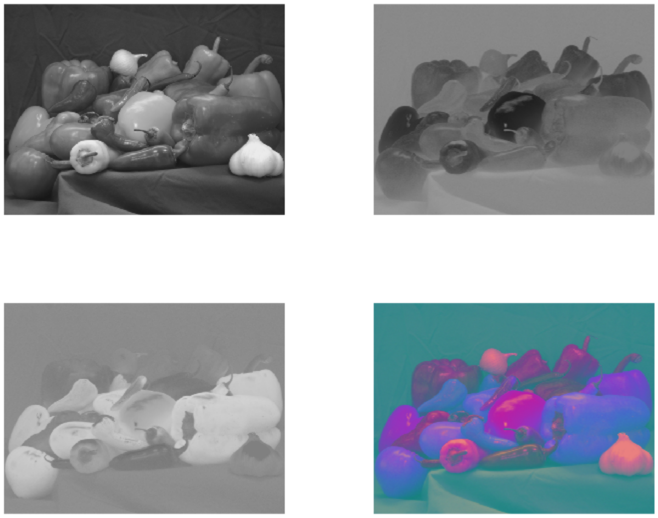

Figure 11: From RGB to the YCbCr and individual channel

3.6 Gray scale color space

The gray scale color space could indicate the intensity (how bright is a particular pixel) of a certain image. The information is represented by a two dimensional array of bytes, which is equal to the width and height of the image. Additionally, a gray scale image has only one channel, which represents the intensity of the white. Luminance is very important in distinguishing visual features. Gray scales transformation is quiet essential if we want to detect features such as edges from a picture.

Source:`https://stackoverflow.com/questions/12752168/why-we-should-use-gray-scale-for-image-processing`

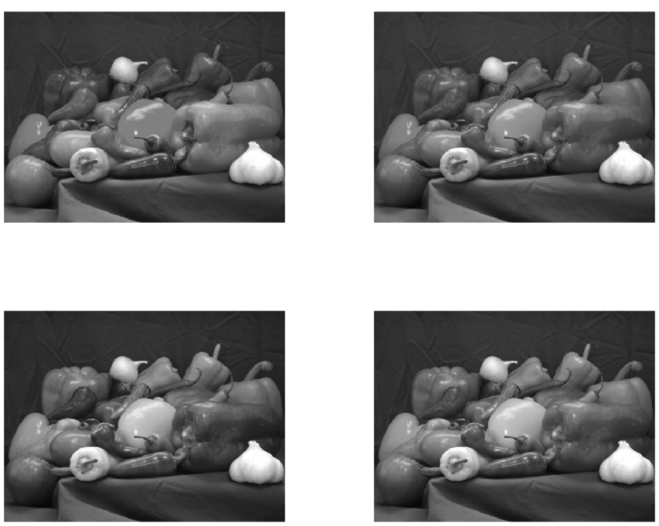

Figure 12: From RGB to the different gray pictures and individual channel

3.7 More on Color Spaces

CMYK color space model: The CMYK color model is the quadruples color space model widely applied in printing. CMYK represents four inks in some color printing, C(cyan), M(magenta), Y(yellow) and K(black). Compared to addictive color model such as RGB, white is the neutral color of the backgrounds, whereas black results from a full combination of different color inks.
CIE Lab color space model: The CIE lab color space produces a color space that is more perceptual linear than other color spaces. More specifically, the same amount of change in each channel will produce the same results. LAB describes mathematically all perceivable colors in three dimensions: L for lightness, a for red-green and b for blue-yellow.The lab color space model is used when graphics for print have to be converted from RGB to CMYK, as the lab gamut includes both the RGB and CMYK gamut. It is more commonly used for surface light rather than mixtures of light.

4 Intrinsic Image Decomposition

4.1 Other Intrinsic Components

The specular component: It could account for the highlights caused by viewpoint, geometry and illumination. More specifically, the specular term could be explained by light rays that reflect directly off the surface, creating visible highlights in the image. Image decomposition is the process of separating an image in conceptually and theoretically different components, under two perspectives 1) addressing content-specific vision applications 2) studying image pattern formation. **Geometric structure component:** The structure part for 1D feature detection, segmentation, object recognition and shape analysis. **Texture component:**The texture component can be used for texture' segregation and classification, surface analysis ,shape/orientation from texture.

4.2 Synthetic Images

The advantages of using synthetic images:

- Shaded images can be generated with different light source directions for the same surface.
- The true depth information enables us to compute the error and compare the performance.

5 Recoloring

1. The true material color can be obtained by analyzing the reflectance picture. We assume that the center point of the picture reveals the pure color of the picture. The middle point returns the RGB of (184, 140.9, 107.99), which corresponds to brown in the color scheme.
2. (Outputs are attached in the project folder)
3. This is caused by the shading effect in the graph. The algorithm ($I(x) = R(X) \times S(X)$) clearly explains the transformation from the pure reflectance to the observed image. The multiplication changes the value of color in the color distribution. Hence, the color on the ball surface is not uniformly distributed.

6 Color Constancy

6.1 Grey world

6.1.1 2

Examples when gray-world algorithms fail

- Under the condition of incandescent lighting on overcast days, the output images might not be satisfactory. It is due to the fact that the gray-world algorithms automatically generate exposure on pictures. It might overcompensate the strong tint and generate a blue picture.
- If we take a picture of green mountains in the foggy weather, the algorithm will cover green trees and grass with reddish color which makes the picture in a weird warm tone.

Reasons: The assumption of gray-world algorithm is that "The average reflectance in a scene under a neutral light source is achromatic". The gray-world algorithm is sensitive to large uniformly colored surfaces, which often leads to the failing of the assumptions.

6.1.2 3

Simpler Lambertian models

- **Max-RGB:** Assumption: "The maximum reflectance in a scene is achromatic" In other words, the surface with perfect reflectance property will reflect the full range of light it captures. Hence, the color of this perfect reflectance is exactly the color of the source.
- **Gray-edge:** Assumption: "The pth Minkowski norm of the image derivative is achromatic." The gray-edge methods are based on the spatial derivation of order n. The Minkowski-norm p that determines the relative weights of the multiple measurements from which the final illumination is estimated. A high Minkowski norm emphasizes larger measurements, whereas a low Minkowski norm equally distributes weights among the measurements.

Name	Symbol	Equation	Hypothesis
Gray-world	$\mathbf{e}^{0,1,0}$	$\left(\int f^c(\mathbf{x})d\mathbf{x}\right) = ke^c$	The average reflectance in a scene is achromatic
Max-RGB	$\mathbf{e}^{0,\infty,0}$	$\left(\int \lvert f^c(\mathbf{x})\rvert^\infty d\mathbf{x}\right)^{\frac{1}{\infty}} = ke^c$	The maximum reflectance in a scene is achromatic
Shades of gray	$\mathbf{e}^{0,p,0}$	$\left(\int \lvert f^c(\mathbf{x})\rvert^p d\mathbf{x}\right)^{\frac{1}{p}} = ke^c$	The pth Minkowski norm of a scene is achromatic
General gray-world	$\mathbf{e}^{0,p,\sigma}$	$\left(\int \lvert (f^c)^\sigma(\mathbf{x})\rvert^p d\mathbf{x}\right)^{\frac{1}{p}} = ke^c$	The pth Minkowski norm is achromatic after local smoothing
Gray-edge	$\mathbf{e}^{1,p,\sigma}$	$\left(\int \lvert (f^c)^\sigma_{\mathbf{x}}(\mathbf{x})\rvert^p d\mathbf{x}\right)^{\frac{1}{p}} = ke^c$	The pth Minkowski norm of the image derivative is achromatic
Max edge	$\mathbf{e}^{1,\infty,\sigma}$	$\left(\int \lvert (f^c)^\sigma_{\mathbf{x}}(\mathbf{x})\rvert^\infty d\mathbf{x}\right)^{\frac{1}{\infty}} = ke^c$	The maximum reflectance difference in a scene is achromatic
Second-order gray-edge	$\mathbf{e}^{2,p,\sigma}$	$\left(\int \lvert (f^c)^\sigma_{\mathbf{xx}}(\mathbf{x})\rvert^p d\mathbf{x}\right)^{\frac{1}{p}} = ke^c$	The pth Minkowski norm of the Second-order derivative is achromatic

Figure 13: All color consistency models

Physics based method This method implements the dichromatic reflection model of image formation, which makes use of information about the physical interaction between the light source and the objects in a scene. Assumption: All pixels of one surface will fall on the same RGB color space. If multiples of such planes are found, corresponding to various different surfaces, then the color of the light source is estimated using the intersection of those planes.The disadvantage of this method is also obvious, it is hard to retrieve such specular reflections and color clipping might occur.

7 Conclusion

Photometric stereo can give reasonable results even when the number of images available is limited. Considering multiple integration paths for the surface height map is also a good idea. The standard procedure of assuming gray scale images is also easy expendable to RGB-valued images, by assuming 3 channel albedos and converting RGB-values to a gray scale value with some linear combination.
For the color space model, the 5 models could be adopted for different purposes, which could be meaningful if we want to extract certain features from the image. Furthermore, we investigated the method for intrinsic image decomposition. We figured out the reflectance component could indicates the true color of the object.
With respect to color Consistency models, while lambertian and physic based models have their own merits and demerits, it is essential for us to identify them and make the best use of them in the future research.