Study Report

1. *Szeliski's book*

First, I read Chapter 2 in order. The Section 2.1 is hard to understand because I am not familiar with some math terminology. It always takes me half of the day to read one or two pages. For example, I can't understand homogeneous coordinates. Then I googled and finally understood it. But this is a waste of time. I later found a way to go to the library to borrow Chinese books in the direction of computer vision and they help me to understand those math terminology quickly.

2.1.1

I learned how to represent 2D points and 3D points using homogeneous coordinates, as well as 2D lines, 3D planes, 3D lines. I can't really understand 2D conics and 3D quadrics.

2.1.2, 2.1.3

2D transformations use a matrix to achieve. This part is relatively easy for me because I have learned two courses about matrix. The accumulation of various transformations can be expressed by the product of the matrix. The adjoint of the matrix can be directly used to transform a line equation.

2.1.4

3D rotation can be obtained by multiplying twice the normal vector to get the vertical component. And use a series of ope rations to get Rodriguez's formula.

$$\boldsymbol{R}(\hat{\boldsymbol{n}}, \theta) = \boldsymbol{I} + \sin\theta[\hat{\boldsymbol{n}}]_\times + (1 - \cos\theta)[\hat{\boldsymbol{n}}]_\times^2$$

Then also there are some expansions of Rodriguez's formula.

Another way to represent 3D rotation is to use unit quaternions. I don't understand this part clearly. I don't understand what's the meaning of ω in unit quaternions.

2.1.5

This part tell me some projection models.

Orthography: drop the z component

Scaled orthography: multiply the scaling factor on the previous basis

Para-perspective: being projected parallel to the line of sight to the object center and then projected onto the final plane normally

Perspective: drop the ω component

After reading Section 2.1, my reading speed for Section 2.2 was greatly accelerated. When I read the paper later, this section helped me a lot in understanding the mathematical and physical foundations of the previous part of the paper. I learned that images are made up of discrete color or intensity values and learned some models to describe these.

2.2.1

A point light source originates at a single location in space and has an intensity and a color spectrum(a distribution over wavelengths).

Area light sources can be modeled as a finite rectangular area emitting light equally in all directions.

Environment map: more complicated, maps incident light directions $\hat{v}$ to color values.

$$L(\hat{v};\lambda)$$

2.2.2

The most general form: the bidirectional reflectance distribution function(BRDF)

$$f_r(\theta_i, \varphi_i, \theta_r, \varphi_r; \lambda)$$

The BRDF is reciprocal because of the physics of light transport.

We can integrate the product of the incoming light L with the BRDF to calculate the amount of light exiting a surface point in a direction.

$$L_r(\hat{\boldsymbol{v}}_r;\lambda) = \int L_i(\hat{\boldsymbol{v}}_i;\lambda) f_r(\hat{\boldsymbol{v}}_i, \hat{\boldsymbol{v}}_r, \hat{\boldsymbol{n}};\lambda) \cos^+\theta_i \, d\hat{\boldsymbol{v}}_i$$

Or

$$L_r(\hat{\boldsymbol{v}}_r;\lambda) = \sum L_i(\lambda) f_r(\hat{\boldsymbol{v}}_i, \hat{\boldsymbol{v}}_r, \hat{\boldsymbol{n}};\lambda) \cos^+\theta_i$$

Diffuse reflection: the diffuse component scatters light uniformly in all directions. The BRDF is constant.

$$f_r(\theta_i, \varphi_i, \theta_r, \varphi_r; \lambda) = f_d(\lambda)$$

Specular reflection: Incident light rays are reflected in a direction that is rotated by 180° around the surface normal.

$$\hat{s}_i = v_{\parallel} - v_{\perp} = (2\hat{n}\hat{n}^T - I)v_i$$

And the amount of light reflected depends on the angle of s and v.

Phong shading combined the diffuse and specular components of reflection with ambient illumination that objects are generally illuminated not only by point light sources but also by a general diffuse illumination corresponding to inter-reflection or distant sources.

The ambient term depends on the color of both ambient illumination and the object:

$$f_a(\lambda) = k_a(\lambda) L_a(\lambda)$$

Phone shading model:

$$L_r(\hat{\boldsymbol{v}}_r;\lambda) = k_a(\lambda) L_a(\lambda) + k_d(\lambda) \sum_i L_i(\lambda) [\hat{\boldsymbol{v}}_i \cdot \hat{\boldsymbol{n}}]^+ + k_s(\lambda) \sum_i L_i(\lambda) (\hat{\boldsymbol{v}}_r \cdot \hat{\boldsymbol{s}}_i)^{k_e}$$

2.2.3

Basic lens model

Object distance, image distance and focal length have a certain relationship.

Chromatic aberration: light of different colors will focus at slightly different distances. We can use models or functions to compensate them.

Vignetting: the tendency for the brightness of the image to fall off towards the edge of the image.

I read Section 2.3 quickly. I am quite familiar with Sampling and aliasing(2.3.1) which is my major, such as Nyquist frequency and Fourier transform.

Then I read 2.3.2(color) and 2.3.3(compression) like reading a novel. They are very interesting but beyond my understanding. I just get familiar with some basic conceptions.

Similarly, I quickly read Chapter 12.

2. Counterfeit Detection Using Paper PUF and Mobile Cameras

2.1 The main work of this paper

Prior work showing high matching accuracy:

1. Using a consumer-level scanner for estimating a projected normal vector field of the surface of the paper
2. Using an industrial camera with controlled lighting to obtain an appearance image of the surface

Limitation:

1. The uncontrolled nature of the ambient light

Work in the paper:

1. Using multiple camera-captured images of different viewpoints to estimate the microscopic normal vector field of paper

Advantage:

1. Relax the authentication setups

Requirements:

1. The mobile captured images should be comparable in resolution and contrast to those captured by scanners
2. Controlling the light(using the flash next to the camera)

2.2 Mathematical and physical principles and basic models

1. Light reflection models

The paper assume that the majority of locations follow the fully diffuse reflection model and the remaining locations are outlier under this model.

$$l_r = \lambda \cdot l \cdot \boldsymbol{n}^T \boldsymbol{v}_i$$

2. Patch registration

My understanding of the setups:

First, use a Hough transform(I can't understand that although I have searched online) to align.

Then, compensation for perspective transformation based on a circle, because the circle is a known reference pattern(I am not sure if I should understand it as such, and I don't know the specific calculation process of perspective transformation).

Finally, use the perspective transform matrix to get the lens location and the direction of incident light.

3. Authentication test

2.3 Paper authentication

1. Norm maps by scanners

Use images scanned from 4 different orientations of the paper and cancel the effect of the unknown incident direction of the scanner light by using two scans obtained in opposite directions. The results show a good authentication performance.

2. Appearance images by industrial cameras

Use two industrial cameras at a high elevation with a semi-controlled lighting condition(a fixed circular ring-shaped light source). The results show it has a good authentication performance. However, if tested under uncontrolled light, the performance will be worse.

3. Using flashlight

Use build-in cameras of 3 different mobile cameras with or without flashlight to get the images of patches in different light-condition rooms. The experimental results show the authentication performance is significantly improved with the help of a flash. And ambient lighting conditions do not have an important role if there is a flash.

4. Proposed authentication

The paper think there exists a gentle spatial intensity change at large scale with circular-shaped level curves. So the macroscopic intensity should be compensated. The author approximates the macroscopic intensity by the averaged perceived intensity of background pixels over a small neighborhood around a pixel location.

3. *Enhanced Geometric Reflection Models for Paper Surface Based Authentication*

3.1 The main work of this paper

Examine several candidate mathematical models for camera captured images of paper surfaces and compare the modeling accuracies with reference to the measurement by the confocal microscopy. New mathematical models takes into account both the light reflection and the image acquisition procession to estimate the normal field and reconstruct 3D microscopic paper surfaces from normal vector fields and decompose the surfaces into different spatial-frequency subbands.

Conclusion:

1. The model with distinct intensity bias for images captured from different viewpoints can provide the closest result to confocal measurement.
2. High-frequency subbands of reconstructed 3D surfaces are more powerful than the norm map in describing the uniqueness of a physical surface.

3.2 Mathematical and physical principles and basic models

1. Diffuse reflection model

The model and compensation for the effect of the macroscopic intensity is same as the previous paper.

2. 3D reconstruction

The intuitive approach to reconstruct a 3D surface is to spatially integrate the gradient field. However, because of discontinuities of the surface, the gradient may be non-integrable. In the paper, the surface is reconstructed by a method named shapelet.

This part is too short so I can't understand this method. I just have a vague idea of this.

3. Difference-of-Gaussian representation

It can decompose the reconstructed 3D map into different spatial-frequency subbands to investigate the surface trend and the spatial changes of the surface.

3.3 Experimental progress and results

1. Proposed models for estimating normal

The author proposed models with intercepts in addition to the diffuse reflection component.

$$y^{(k)}(p) = \lambda l^{(k)}(p) \cdot n(p)^T v(p) + {\beta_0}^{(k)}(p)$$

A macroscopic intensity field that contain the global factors(the strength of the arriving light, the overall intensity bias, the effect of the camera direction)

$$\tilde{y}^{(k)}(p) \approx E\left[y^{(k)}(p)\right] = \lambda l^{(k)}(p) \cdot m_z v_z(p) + {\beta_0}^{(k)}(p)$$

The paper got the estimates for $\lambda l^{(k)}(p)$ and ${\beta_0}^{(k)}(p)$ using least-squares. Then it assumed the bias in intensity has the same effect for all images so the normal vector at each location could be estimated.

2. Experimental results

The two reconstructed 3D surface appears differently at the patch scale, but after decomposing the surface into different spatial-frequency subbands, subbands of high spatial-frequencies have many more overlapped peaks and valleys than the full spectrum curves from the original surface. Also, they have better discrimination capability.

3. Practical authentication system

This section introduces a practical application.

4. *EXPLORING NONTRIVIAL RELATIONS USING NEURAL NETWORKS FOR PAPER SURFACE BASED AUTHENTICATION*

4.1 The main work of this paper

Previous work:

Using scanner or camera-acquired images based on the assumption that the paper surface reflects the light in a fully diffuse manner to estimate microscopic surface normals

Work in the paper:

Using neural networks which take the specular reflection into consideration to obtain more precise surface normal estimates and reveal seemingly complex relations.

4.2 Mathematical and physical principles and basic models

1. Generalized light reflection model

The perceived intensity l^r at a paper surface location can be modeled as follows:

$$l^r = \frac{l}{\|o-p\|^2}\{w_d \cdot (n^T v)^+ + w_s \cdot ({v_c}^T v_r)^{k_e}\}$$

2. Photometric stereo

Using the scanner to acquire the paper in two opposite directions. The difference has been shown to be in proportional to the y-component of the norm map.

$$\delta I = l \cdot w_d \int_{-a}^{a} n^T \frac{\left(O_x, O_y, O_z\right) - (O_x, -O_y, O_z)}{\left\|(O_x, O_y, O_z)\right\|^3} dO_x = n_y \cdot s$$

4.3 Neural network

I am still reading this part.