Sequential Estimation of 3D Shape and Reflectance Property of Object with Light Field Camera

Yuka Kometani* and Takafumi Koike**

Graduate School of Computer and Information Sciences, HOSEI University,
Kajino 3-7-2, Koganei, Tokyo 184-8584, Japan

*yuka.kometani.6m@stu.hosei.ac.jp, **takafumi@hosei.ac.jp

Abstract

We discuss a method which sequentially estimates 3D shape and reflectance property of an object. To estimate these values, we use a light field camera to capture multi-view light fields. First, we acquire a 3D point cloud by structure from motion, and then we generate a mesh from the point cloud. Second, we simultaneously estimate normal vectors and a light source from the mesh. Finally, we estimate reflectance parameters of a reflection model with the mesh, normal vectors, and the light source. We also compare the original photo with a CG image rendered using the parameters estimated by our method.

Keywords: Structure from Motion, BRDF, Light field camera.

1. Introduction

3D CG is rendered using three elements: 3D shape of object, reflectance property of object, and illumination distribution. If we can estimate three elements of real world's objects accurately, we can render the object with high synthesis accuracy. In this research, we focus on the 3D shape and reflectance property which are features dependent on the object.

Structure from Motion (SfM) is a method which estimates the shape of an object and camera pose simultaneously. SfM estimates them by searches same points, called by corresponding points, in the 3D space from feature points each of target images. In order to restore an object as a dense point cloud by SfM, it is necessary to increase the number of corresponding points. Most papers accomplish it by increasing the number of photographing viewpoints. On the other hand, the Bidirectional Reflectance Distribution Function (BRDF) representing reflectance property is a four-

dimensional function that outputs the object reflectance with the incident/reflected direction of the light source as an input. To estimate the BRDF of an object, it requires a huge number of measurements to measure the reflected light, since the result varies according to the light source and the observation direction.

In this paper, we present a method to efficiently estimate the 3D shape and reflectance property of an object using a light field camera. The light field camera decomposes information acquired by the imaging sensor into rays and records them. By extracting a part of the obtained rays, it is possible to generate multiple sub-aperture images with slightly different viewpoints by single shot. Also, since the viewpoint is few changes between sub-aperture images, it is easy to estimate the reflection change in BRDF. Therefore, it is possible to reduce the number of measurements of each estimation using a light field camera. Before this paper, we estimated 3D shape and reflectance property sequentially⁽¹⁾. At that time, there were problems that reflected light we acquired is non-uniform incident/reflected direction and we could not acquire gloss parameter in reflectance property estimation. Therefore, we improved calibration accuracy and subdivided of meshes in this method to improve the reflectance property.

2. Light Field

A light field is a concept representing a field of rays flying through space. The 4D light field used in this research is an expression method of light presented by Levoy et al.⁽²⁾. To represent rays with 4D light field coordinates, we use four parameters at the intersection of the two planes that you want to represent in space, such as the figure 1. A camera, which records 4D light field, is called a light field camera. Light field cameras, incorporated multiple microlenses,

decompose 2D light into 4D rays and acquire by the image sensor with microlenses. After that, cameras record the result as a RAW image, and complement the RGB information between pixels to make a color image (demosaic) as figure 2 (a). The image is a honeycomb structure as shown in figure 2 (b) which is enlarged figure 2 (a). Each hexagon corresponds to one microlens, and it is possible to acquire a set of 4D light field coordinates (u, v, s, t) by the intersection of the ray in the main lens (u, v) and the intersection of the ray in the microlens (s, t) . From the obtained light field, we can acquire 2D images by extracting only the ray set passing through the position (s, t) of the microlens. We call an image with slightly different viewpoints from this intersection (s, t) a sub-aperture image.

Fig. 1. Light field coordinate.

(a) Color image

(b) Enlarged image

Fig. 2. Light field image.

3. Related Work

Snaveely et al. have presented a SfM method which estimates camera pose using bundle adjustment and reconstructs the target object with high accuracy⁽³⁾ to handle large scale image groups. Wu et al. have speeded up the method of Snaveely et al. by omitting calculations performed between images with few corresponding points⁽⁴⁾. However, both Snaveely's and Wu's methods are for reconstructing objects with diffuse reflections, so the accuracy of shape estimation for glossy objects is low. Also, if the number of shots is small, we can obtain only sparse 3D point cloud.

Most BRDF estimation methods use only flat plates as target object. When we use a 3D object to estimate, it is necessary to obtain the relationship between camera poses accurately. Fukuda et al. have implemented a method which estimates reflection parameters of a 3D object consisting of multiple materials by accurately estimating camera pose⁽⁵⁾. However, this method requires to take shots separately for shape and reflectance property estimation.

Wang et al. have presented a method to estimate the depth and reflectance property of glossy objects using a light field camera⁽⁶⁾ simultaneously. The method cannot estimate the 3D shape of an entire object since only one shot is used for the reconstruction. Xia et al. have estimated the shape and reflectance property of an entire object consisting of multiple materials under unknown illumination⁽⁷⁾. Xia's method obtains the shape and reflectance property of a 3D object using a voxel based and surface geometric based method. This method can estimate simultaneously with high accuracy for metallic objects which were low accuracy estimation of conventional method. However, this method has limitation

that the number of shots increases by taking a video of the target object with rotating and moving.

In our method, we use Wu's method to estimate the 3D shape of a single material object, and estimate the reflectance property from the obtained shape. Unlike Fukuda et al.'s method, our method cannot work with objects consisting of multiple materials. Instead it is able to decrease the number of shots and to estimate 3D shapes and reflectance property at once.

4. Method Overview

We use a group of still images taken with a light field camera to sequentially estimate the 3D shape and reflectance property of an object. Figure 3 shows a procedure of the method. We describe details of each operations below.

- **Preprocessing**

We need to calibrate the light field camera to correct images of target object. We photograph of a checkerboard from multiple viewpoints and search the coordinates of the grid points in each image. We acquire the internal parameters of the camera by comparing the coordinates of the grid points for each image.

- **Acquisition of input data**

We acquire pictures of an object and enhance these images. We photograph of a target object from various directions with the light field camera, then we estimate the

Fig. 3. Estimation procedure.

focal length and synthesizes light field from the photographed data for each viewpoint. We enhance the color tone of image using the gamma value and the white balance coefficient obtained from the photographed data. Also we enhance the distortion of the image on the basis of the parameter obtained in the preprocessing.

• Estimating 3D shape

We estimate 3D point cloud by SfM, and obtain 3D shape. At first, we collect sub-aperture images from each still image to generate an image group. We obtain sets of 2D corresponding points in each image, which represent the same point in 3D space. We can obtain each corresponding point from executing feature point matching with SIFT to pairs of images extracted from the image group. We compare the coordinates of the acquired corresponding points to estimate the relative pose between the two viewpoints. After that, we obtain the 3D point cloud representing the target object by estimating the 3D coordinates of these corresponding points from the relative pose between each pair of images from the image group.

We reconstruct the 3D shape of the object by applying a mesh to 3D point cloud (Making face). We make face by the Ball Pivoting algorithm which estimates the surface by rolling the sphere on the point cloud⁽⁸⁾. After that, we subdivide surfaces to add points along each faces in mesh.

• Estimating reflectance property

Estimating the BRDF of the object needs the incident/reflected direction of light obtained from the 3D

shape of the object and relative pose of the camera which receive reflected light. For each image, we estimate the reflected direction from the viewpoint obtained from the camera pose acquired during shape estimation. We estimate the incident direction as the normal vector of the 3D point with the brightest reflected light.

4.1 Estimating incident/reflected direction

We estimate the BRDF of the object based on information obtained by 3D shape estimation. First, we acquire the reflected direction. The ray reflected at a certain point X on the object surface passes through its corresponding point in each image and heads toward the image's viewpoint. Therefore, we can obtain the reflected direction by calculating the vector from the 3D point to the camera viewpoint V_{i0} is in the 3D space. It uses the pixel value of the corresponding point which V_{i0} passed through to represent the color of the reflected light. We obtain the 3D coordinates of a viewpoint by transforming its position in camera coordinates $(0, 0, 0)^t$ into the 3D coordinate of the point cloud. Thus, we solve for the value of V_{i0} using equation (1).

$$\frac{1}{z} \begin{pmatrix} x_i \\ y_i \\ 1 \end{pmatrix} = K_i \left(R_i \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} + t_i \right) \quad (1)$$

Here, $(X, Y, Z)^t$ is the 3D coordinates of point X , $(x_i, y_i)^t$ is the image coordinates of the corresponding point x in image i , and R_i , t_i is the rotation matrix and translation vector representing the camera pose. K_i is defined by the following matrix using f_i , the focal length of the camera:

$$K_i = \begin{pmatrix} f_i & 0 & 0 \\ 0 & f_i & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (2)$$

Substituting $V_{i0} = (X_0, Y_0, Z_0)^t$, the coordinates of the viewpoint in its image, into equation (1) results in the next equation:

$$0 = K_i(R_i V_{i0} + t_i). \quad (3)$$

Since K_i is a diagonal matrix, its inverse matrix always exists. Also, the inverse of rotation matrix R_i can be obtained by inverse transformation. Using equation (3) and these two inverse matrices, the 3D coordinates of the viewpoint is as follows:

$$V_{i0} = -R_i^{-1} t_i. \quad (4)$$

From this V_{i0} , the reflected direction V can express as $V = V_{i0} - X$.

Next, we acquire the incident direction. The information to estimate incident direction is different by the

type of light source. We assume light source to be a single area light source, then we acquire the incident direction as inverse vector of the direction of light source. We use the normal direction at a brightest point as the incident direction. Here, brightest point is the point on 3D shape that the average of pixel values of corresponding points is brightest among 3D points. We estimate a reflectance parameters of reflection model described in the next section using the incident/reflected direction obtained in this section.

4.2 Reflection model

The color of reflected light on the object surface is represented by the sum of diffuse color and specular color. We estimate the reflectance property of object as the parameter of reflection model. We use Lambert reflection model as diffuse reflection and Blinn-Phong reflection model as specular reflection. The luminance L_r for each color component (in this study, RGB as a component) represents by the following equation (5) by the expression of reflection model.

$$L_r(\mathbf{L}, \mathbf{V}, \mathbf{N}) = C_d(\mathbf{L} \cdot \mathbf{N}) + C_s((\mathbf{H} \cdot \mathbf{N})^m) \quad (5)$$

C_d , C_s , and m represent the diffuse reflectance, the specular reflectance, and the glossiness of the object's material, respectively. These are constants that differ for each kind of material. $\mathbf{L}$ and $\mathbf{N}$ each denote the incident direction and normal. $\mathbf{H}$ is a half vector defined as $\mathbf{H} = \frac{\mathbf{L} + \mathbf{V}}{|\mathbf{L} + \mathbf{V}|}$.

When the parameter of reflection model we estimated matches the reflectance property of an object, pixel value c at an arbitrary point on the object surface matches the intensity of reflected light L_r calculated from the incident/reflected angle and normal at the point. Based on this fact, we estimate the reflectance property of objects by obtaining the value of C_d , C_s , and m which minimizes the following equation (6).

$$F = L_r(\mathbf{L}, \mathbf{V}, \mathbf{N}) - c \quad (6)$$

Since F is nonlinear function, we estimate each color components by the Levenberg-Marquardt method (LM method) for solving the nonlinear least squares problem.

5. Implementation

We have used Lytro Desktop 3.1.1 and MATLAB R2016a on a Windows 10 PC (CPU: Intel Core i7-6700, Memory: 16.0GB) to implement our method. The light fields we captured by the first generation Lytro light field camera. We also uses MeshLab v1.3.3 to confirm the acquired point

Fig. 4. Target object.

cloud. We calibrate camera by Dansereau et al.'s method with a 19 x 19 checkerboard⁽⁹⁾. We also use VisualSfM⁽¹⁰⁾ for shape estimation.

6. Experiment

We visually compared real objects with the CG objects reconstructed from the objects using the results of the 3D shape and reflectance property estimation, in order to confirm that it is possible to sequentially estimate object in our method.

6.1 Object and environment

We have experimented using a glossy elephant model with paper clay (roughly $5 \times 3 \times 5$ cm) as shown in figure 4. We took 88 shots of this model while moving the light field camera 360 degrees around the model from some different heights; same or slightly above the model, and extract nine sub-aperture images from each shot. We used a fluorescent lamp, an area light source, in the experiment environment.

6.2 Result

Figure 5 shows on experimental result of the shape estimation. The model included roughly 1000 points, and we could not acquire the 3D shape of the entire object. The estimation results of reflectance property was $C_d = (0.31, 0.21, 0.14)$, $C_s = (0.39, 0.35, 0.36)$, and $m = (0.00, 0.00, 0.00)$. Figure 6 shows the visualization result of measured reflectance property and its estimation result.

Figure 6(a) shows pixel values recorded for each incident/reflected angle. Figure 6(b) shows estimated values from reflectance parameter. We define $c_{\text{error}} = \text{mean}(L_r(\mathbf{L}, \mathbf{V}, \mathbf{N}) - c)$ as error indicator and c_{error} is 0.076. This means each color component of pixel has about 0.076 error. The result was darker and decreased gloss comparing with our previous study⁽¹⁾. In the other hand, we could acquire rays which are non-uniform incident/reflected direction. However, rays with small angle could not acquire, so estimation results remained low in gloss. We reproduced 2D image by applying the estimated shape and reflectance

Fig. 5. Result of shape estimation.

Fig. 6. Comparison between real data and estimation.

property to the visualization of the reconstructed mesh with estimated color. Figure 7 shows comparison between the original model (Figure 7(a)) and the reproduced 3D model (Figure 7(b)). Comparing the original image with the reproduced image, we can see that we obtained an output similar to the original image. However, since the glossiness m is small and the influence of specular reflection is small, it could not say accurate estimation to have been made.

7. Discussion

It seems that the cause of low accuracy of shape estimation is to integrate partial shapes with low accuracy. We use multiple sub-aperture images. Light field camera acquires 120 images decomposed from input information to the sensor, total pixels is almost the same amount by a general camera. As a result, the number of pixels decreases in single image. Furthermore, the pose change between sub-aperture images obtained in single shot is as small as each coordinates in images of corresponding points move only by less than one pixel. Therefore, we think that our method could not estimate accurately the relative viewpoint pose of sub-aperture images. We think that we should change a SfM method to use images with more pixels or comparing rays rather than points for improve the accuracy of estimating relative pose between sub-aperture images.

Fig. 7. Comparison between original and CG

Regarding low accuracy of reflectance property estimation, we thought it was due to an error of camera pose and lack of partial shape with strong gloss, so we added operations this time. About the error of the camera pose also mentioned in the shape estimation. We think that we could get correspondence between the point on the image with high gloss and corresponding 3D points of the target object hardly because of the error of the pose.

About the lack of partial shape with gloss, we subdivide meshes to close holes in this method because the information of the gloss decreases if there is a hole. Compared with the result of previous method⁽¹⁾, non-uniform incident/reflected direction of rays decreased. However, we could not improve the estimation accuracy because we could increase few rays we obtained with high luminance. We think that this is due to the error of the normal caused by the error of shape estimation. Rays with small angle of incident/reflected direction are fewer than other rays and hard to acquire, so we think it was susceptible to error of the normal. Therefore, we have to improve the accuracy of shape estimation to perform normal estimation based on mesh estimation, in order to improve the accuracy of the reflectance property estimation. Similarly we have to acquire rays with small angle to make it less susceptible to errors.

8. Conclusions

We have presented a method for sequentially estimating the 3D shape and reflectance property of an object using a light field camera. Experimental results show our method can obtain a rough partial shape of the object. On the other hand, the estimation error of SfM increased using the sub-aperture image, and it turned out that making mesh and normal estimation cannot be performed accurately. We confirmed that we can estimates reflectance property even by using the same image group as shape estimation. However, we found that lack of partial shape with gloss and errors of 3D shape estimation. They are resulting from using sub-

aperture image similar to the estimation by a general camera.

In the future, we will aim at implementing estimation of object with high accuracy by improving the shape estimation with a method using rays such as Wang et al.'s method⁽⁶⁾. Also we will aim at reducing errors each of estimation by improving the details of the method such as accurately calibration, exclusion of outliers in reflectance property estimation. In addition, we will estimate with changed light irradiation method, and attempt to improve the accuracy of reflectance property estimation in order to prevent lack of partial shape cause by gloss.

- (10) Cangchang Wu, "VisualSfM: A Visual Structure from Motion System", <http://ccwu.me/vsfm/> (2016/12/26).

References

- (1) Yuka Kometani and Takafumi Koike, "Sequential Estimation of 3D Shape and Reflectance Property of Object with Light Field Camera," VC/GCAD Joint Symposium, 2017. (in Japanese)
- (2) Marc Levoy and Pat Hanrahan, "Light field rendering," Proc. ACM SIGGRAPH, pp.31-42, 1996.
- (3) Noah Snavely, Steven M. Seitz, and Richard Szeliski, "Photo Tourism: Exploring Photo Collections in 3D," Proc. ACM SIGGRAPH, pp.835-846, 2006.
- (4) Cangchang Wu, "Towards Linear-time Incremental Structure from Motion," Proc. 2013 International Conference on 3DTV, pp.127-134, 2013.
- (5) Hisato Fukuda, Hiroshi Kawasaki, Yoshinori Kuno, and Ryo Furukawa, "Efficient reflectance property estimation of 3D object's surface", The Special Interest Group Technical Reports of IPSJ, 2010-CVIM-172, No.8, pp.1-8, 2010. (in Japanese)
- (6) Ting-Chun Wang, Manmohan Chandraker, Alexei A. Efros, and Ravi Ramamoorthi, "SVBRDF-Invariant Shape and Reflectance Estimation from Light-Field Cameras," IEEE Conference on CVPR, pp.5451-5459, 2016.
- (7) Rui Xia, Yue Dong, Pieter Peers, and Xin Tong, "Recovering Shape and Spatially-Varying Surface Reflectance under Unknown Illumination," ACM TOG, Vol.35, No.6, p187, 2016.
- (8) Fausto Bernardini, Joshua Mittleman, Holly Rushmeier, Cláudio Silva, and Gabriel Taubin, "The Ball-Pivoting Algorithm for Surface Reconstruction," IEEE TVCG, Vol.5, No.4, pp.349-359, 1999.
- (9) Donald G. Dansereau, Oscar Pizarro, and Stefan B. Williams "Decoding, Calibration and Rectification for lenselet-based plenoptic cameras," IEEE Conference on CVPR, pp.1027-1034, 2013.