# 3D reconstruction of image from Vision Based Tactile Sensor

Won Kyung Do
Stanford University
440 Escondido Mall, Stanford, CA 94305
wkdo@stanford.edu

## Abstract

*Vision-based tactile sensor has been widely used since the robot requires rich feedback for great performance on manipulation task. The tactile sensor with fisheye lens can achieve small form factor with wide field of view. For such sensor, the 3D reconstruction of the surface of gel from the single image is critical problem. In this paper, I proposed three different approach for 3D reconstruction of the tactile sensor's surface: 1)end-to-end network approach from 3D pose of the object, 2)reconstruction with 3D pointcloud, and 3) 3D reconstruction from the 3D-printed indicator with least square method by using nonlinear conversion matrix for image-3D position.*

## 1. Introduction

Tactile sensing has been contributing to increase robot dexterity. Especially, rich tactile feedback from sensor is essential for robust object manipulation and control. Recently, vision-based tactile sensors have been outperformed previous sensors due to high resolution. High resolution vision based sensors provides a solution to this problem by detecting both shear and normal forces [11], controlling grasping motion [17], or adjusting grasp [8].

However, most vision-based tactile sensors are expensive, bulky, and are limited on the 2D shape. They shares the similar design approach: most vision-based sensors observe a image of the deformed surface through the clear gel with camera. Vision-based tactile sensor such as Gelsight, Gelslim, or DIGIT has high resolution because of camera but flat surface limits the application in manipulation tasks [6, 7, 10]. Therefore, the tactile sensors with 3D curved surface design is essential for various robot manipulations.

Camera with fisheye lens can be one of the solution for designing 3D-curved tactile sensor for robot manipulation. Fisheye lens with wide field of view(FOV) allows to detect 3D surface more efficiently. From our lab, we proposed the vision-based tactile senor with fisheye lens 1. Monocular camera with mounted gel enables to get the area contact

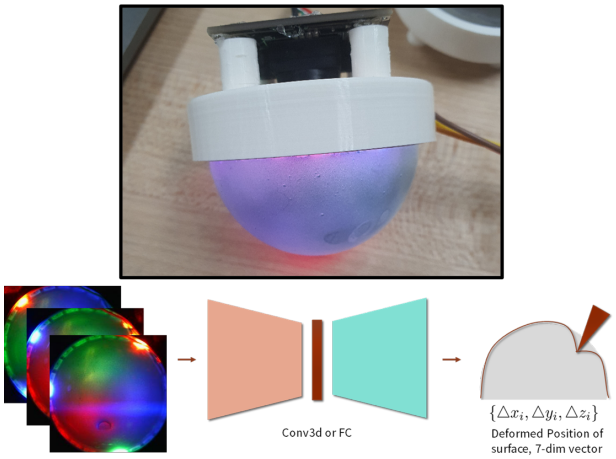


Figure 1. The tactile sensor with fisheye lens(upper image) and one of the proposed approach for 3D reconstruction of sensor surface(lower image).

from soft surface of the gel, which is covered with reflective surface. With the right illumination on the sensor the tactile feedback is available with single image. From the high-resolution image, we can extract corresponding 3D surface of the sensor. However, major challenge with the fisheye lens is 3D reconstruction of the surface from the 2D image. Scaramuzza proposed 2d calibration for omnidirectional camera [16]. Indeed we can reconstruct 3D surface from the fisheye lens from the 2d calibration and proposed 3D reconstruction method from the class. However, We can reconstruct 3D surface of the sensor from the fisheye image directly without using traditional pinhole camera model or Scaramuzza's model. Furthermore, we can fix the surface of the tactile sensor and align the center of the hemisphirical-shaped sensor with the fisheye lens. These constraints will allow a new condition for 3D reconstruction and possibly suggest a more precise result than using traditional technique.

Therefore, this project aims to **reconstruct a 3D surface of the tactile sensor with fisheye lens camera** without using traditional 3D reconstruction way.

## 2. Related Works

## 3. Problem statement

The main focus on this project is 3D reconstruction of the tactile surface from single image input. First, we have to collect the valuable dataset for image-depth pair from given sensor. Next, based on the collected dataset, I am aiming to reconstruct the 3D tactile sensor surface.

Tactile sensing has been contributing to increase robot dexterity. Especially, rich tactile feedback from sensor is essential for robust object manipulation and control. Variety of tactile sensors has been developed in many different approach such as piezoelectricity [18], optics [6], resistance [5], or capacity [9]. Recently, vision-based tactile sensors have been outperformed previous sensors due to high resolution. High resolution vision based sensors provides a solution to this problem by detecting both shear and normal forces [11], controlling grasping motion [17], or adjusting grasp [8]. However, most vision-based tactile sensors are expensive, bulky, and are limited on the 2D shape. Vision-based tactile sensor such as Gelsight, Gelslim, or DIGIT has high resolution but flat surface limits the application in manipulation task [6, 7, 10]. The tactile sensors with 3D curved surface design such as Omnitact [12] allows multi-directional sensing but too expensive to use on real application. Round fingertip sensors from [14] also has 3D shape, but relatively small resolution comparing to other vision-based sensors. Therefore, the tactile sensor( 1) with fisheye lens is able to ensure the high resolution, and detecting geometrically wide range.

The 3D reconstruction from single image have been widely proposed [15]. A quadratic model for Lambertian reflectance under arbitrary illumination conditions, which is the property that our sensor has, could be helpful to reconstruct the 3D surface of the tactile sensor [2]. Another possible approach for 3D reconstruction for tactile sensor can be the end-to-end reconstruction [10]. This approach has been widely used for vision-based tactile sensor with acceptable accuracy. Scaramuzza proposed 2d calibration for omnidirectional camera [16]. This approach is not useful for 3D reconstruction of gel surface, however, the omnidirectional camera with fisheye lens can also be parameterized like pinhole camera with this appropach.

## 4. Technical Approach

I proposed three different approach to reconstruct the 3D surface of the gel surface from single image. The object of three approaches are: reconstruction of 3D surface from image with using different property of the tactile sensor.

### 4.1. 3D reconstruction with end-to-end approach

The first approach for the 3D reconstruction is directly linking between image from tactile sensor with the 3D pose of the detected object. Fig. 2 shows the diagram of the approach. If we only use the same object with known shape, the network can analyze the difference from the given 3D pose. The output of network can be defined by

$$\text{Network Output} = \left[x, y, z, q_0, q_1, q_2, q_3\right] \in \mathbb{R}^7 \quad (1)$$

Where $[q_0, q_1, q_2, q_3]$ is quarternion of the poked object. Furtheremore, as depicted in the Fig. 2 the tactile sensor allows to get multiple images from one observation by changing the position of LED with maintaining cyclic symmetry. For the approach, I used 3 symmetric - LED position with different color - Red, Green, and Blue. The image has been cropped and resized into $256 \times 256 \times 3$ size.

Since we are using multiple images(3 images for this case), I used 3D convolutional layer to get the feature. I modulized the network structure with 'block': each block has two continuous series of $(Conv3D, Batchnorm3D, ReLU)$. The network consists of 8 'block's with the size of $(32, 64, 128, 256, 512, 512, 512, 512)$, respectively. Each 'block' has same kernel size 3 and stride as 2. After pass the series of block network, I flattened the network with two linear layer that has input channel as $(165888, 100)$, respectively. Finally, after it pass the ReLU and sigmoid layer, the result has been directly matched with the 3D pose of the poked object. we detect the 3D pose of the poked object from Franka Emika arm with calculated position of end-effector.

### 4.2. 3D reconstruction from 3D pointcloud

Next approach I proposed was 3D reconstruction from the 3D pointcloud. Since the shape of tactile sensor is hemispherical, we have utilize multiple depth sensor to observe the whole pointcloud of the tactile sensor's surface. I obtain the pointcloud from 3 same ToF(Time-of-Flight) sensor, PicoFlexx ToF sensor [13]. Tof sensors are mounted on the 3D printed holder as shown in fig. 4. If we correctly transform and match the orientation of each pointcloud, we can get the 3D pointcloud of tactile sensor's surface. From these pointcloud, we can directly estimate the deflection of the surface of input image. To achieve this, I had to correctly match the pose of the each ToF sensor. After matching the origin of the pointcloud correctly, we can estimate the 3D deflection of the sensor from the pointcloud. We can use similar network approach proposed in above section with different output size.

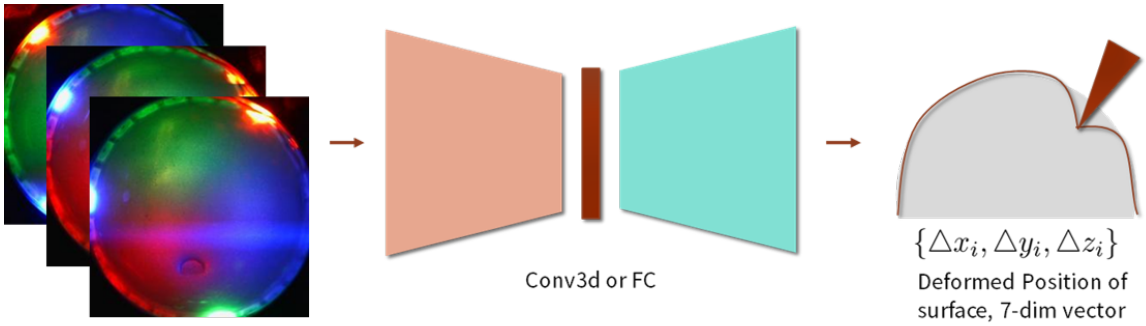


Figure 2. The Network structure of the first approach.

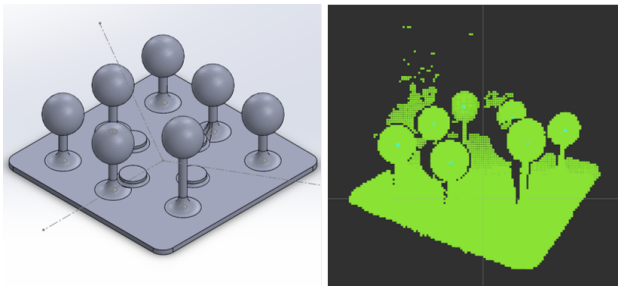

Figure 3. 3D printed part for pointcloud calibration(left image) and one of the pointcloud observed from ToF sensor(right image).

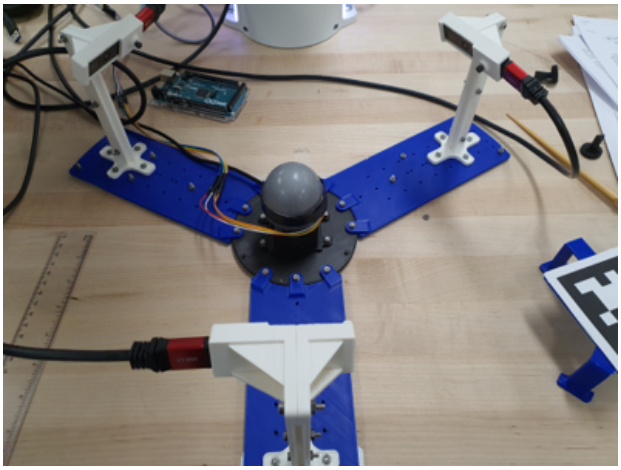

Figure 4. Test environment of tactile sensor with 3 same ToF sensors.

### 4.3. 3D reconstruction from 3D indicator

Camera from fisheye lens have 2D circular image. We can first get the image gradient and use that gradient for designated part. In addition, if we construct a sensor that the light source positioned at infinity, then we can make lambertian surface. Therefore, the fisheye lens calibration can be easier with three constraint: 1) the observed surface will have lambertian reflection, 2) the center of the fisheye lens and the center of the hemisphrical sensor surface align, 3) the position of 3D surface is fixed when the surface of the sensor is not deformed. From these assumptions, the fisheye lens construction can become easier than using general omnidirectional calibration approach from Scaramuzza.

To use above method, I tried to get the information of the sensor surface from 3D printed indicator as shown in Fig. 5. The indicator has hemispherical shape with the same with of the sensor. Therefore, we can directly attach the indicator with the tactile sensor. There are holes with the same size 6mm with equal angular distance. Since we can precisely measure the position of each hole, we can get the information of the hole from the designed CAD file.

Even thought the gel part of the sensor is covered with the reflective surface, we can still observe the lights from outside when we turn off the LED lights installed inside of the sensor. From ambient light source from outside, the tactile sensor can detect the hole with different deflection ratio. By observing deflection ratio of each holes with right position, we can estimate the deflection of each pixel of the input image.

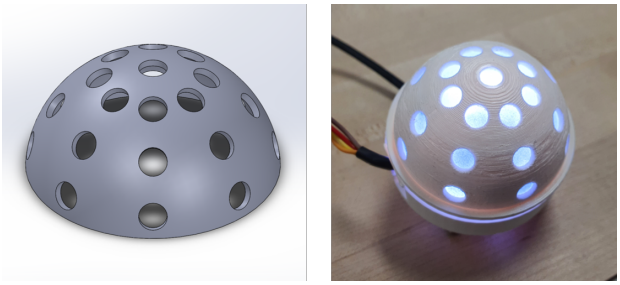

Figure 5. The CAD design of 3D indicator(left image) and the picture of the sensor with 3D indicator attached (right image).

## 5. Experiment Results

### 5.1. 3D reconstruction with end-to-end approach

The dataset for the image-position pair has been collected. The image is corresponded to the object that poked the sensor. Fig.6 shows the example of the image data. However, the data collection step is too slow and the current dataset size is about 1500 image-position pair. The

Figure 6. The image from captured dataset. The object is shown in right-middle part of the image.

train loss and validation loss of the network I proposed was **93.2370, and 57.0529**, respectively.

Apparently, the train/validation loss are huge. Most error comes from the lack of the dataset. Considering the size of the input image and output class, we need to collect the data more than 10,000. However, because of the limitation of the collection step, I couldn't get the appropriate dataset for the training the network. This approach will become useful after I collect the more data from the sensor. The collection procedure of the data requires human input, e.g. the person need to push the each record button while the robot and sensor measures the input image and output pose of the object. If we simplify and automize the collection procedure, I might be more easier to get the dataset.

### 5.2. 3D reconstruction from 3D pointcloud

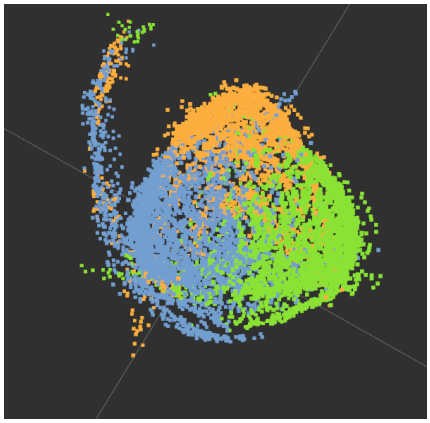

Figure 7. Captured pointcloud data from three different depth sensor.

From the experiment, we can collect corresponding 3D pointcloud dataset for 2D fisheye image. However, the pointcloud data does not have same origin from 3 ToF sensor. Fig. 7 shows raw image with erroneously aligned origin. we can clearly see that the origin of the pointcloud has small bias.

To solve this problem, I designed the 3D object with multiple sphere as shown in Fig. 3. I tried to get the center of sphere from each pointcloud sensor. Since we know the radius of each sphere, we can measure the centers of spheres by subtracting known radius vector from the center of the observed circle from the depth sensor image. There was occlusion problem with some spheres,( you can see in the right image in Fig. 3, the last second sphere from upper left corner makes some occlusion.) so I manually selected the center of sphere for that case.

From given set of the center of sphere in each pointcloud frame, I tried to match those set by using ICP(Iterative closest Point) algorithm [3]. If there is two pointcloud set $(p, q)$ and if we want to find the appropriate rotation matrix $R$ and displacement $t$, we can find

$$p = \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} = R \begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix} + t = Rq + t$$

such that

$$\min_R E = \sum_{i=1}^{n} \|Rp_i + t - q_i\|$$

If we find R and t such that minimizes above error, we can get the appropriate rotation matrix between two pointcloud set. Since I have three different pointcloud set, I tried to get the two rotation matrix $R_{12}, R_{13}$. Fig. 8 shows the result of pointcloud matching by using ICP algorithm. I get the following rotation matrix in homogeneous form:

$$R_{12} = \begin{bmatrix} 1.0000 & 0.0013 & -0.0063 & 0.0071 \\ -0.0011 & 0.9997 & 0.0229 & -0.0163 \\ 0.0063 & -0.0229 & 0.9997 & -0.0038 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$R_{13} = \begin{bmatrix} 0.9987 & -0.0335 & 0.0390 & 0.0095 \\ 0.0328 & 0.9993 & 0.0174 & -0.0034 \\ -0.0396 & -0.0161 & 0.9991 & -0.0048 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

From above result, I could match the pointcloud correctly (see Fig. 9). The next step I will try to do is training the network between 3D pointcloud and the image. However, because of the lack of the GPU resource, there was huge limitation on training the network with such huge pointcloud dataset. Since I was doing this method in parallel with the first approach, the resource for training this network become much smaller. I will continue to train the network if I can secure proper GPU/server for training.


Figure 8. The result of pointcloud matching by using ICP algorithm.

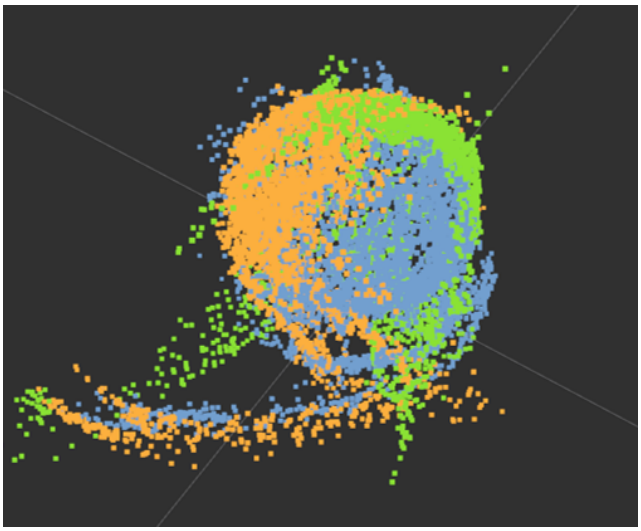

Figure 9. The result of poitcloud after matching the pointcloud set correctly with the ICP algorithm. By comparing the result with the Figure 7, we can notice that the algorithm works well.

### 5.3. 3D reconstruction from 3D indicator

**This part has been accomplished after presentation.**

From the image observed from the tactile sensor attached to the 3D indicator, I tried to get the information of the observed hole from the picture. First, I get the edge and gradient of the image by calculating local gradient of image with Canny method [4]. Then from the detected edge, I calculated the size and the orientation of the ellipse from the image by using Hough Circle Transform [1]. From these procedure, I could get the locus of the hole observed from the fisheye lens camera as shown in the Fig. 10.

Since we already know the corresponding 3D coordinate of these detected holes from the CAD file, we can directly match the 3D position of the tactile sensor with the given

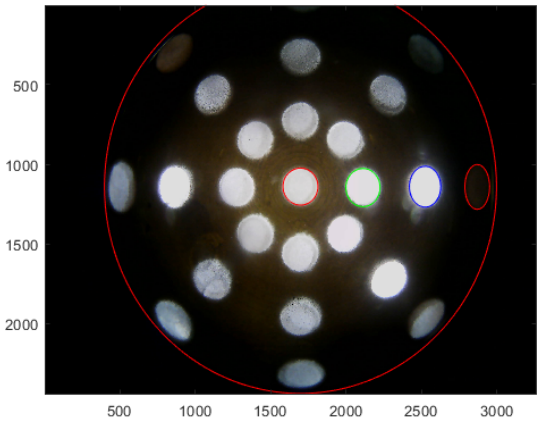


Figure 10. The result of hough transform for ellipse detection on the image from tactile sensor with 3D indicator.

locus of the ellipse. If we set the locus set from the image and corresponding 3D position as

$$pt_{img} = \begin{bmatrix} u_1 - u_c & v_1 - v_c \\ u_2 - u_c & v_2 - v_c \\ \dots & \\ u_n - u_c & v_n - v_c \end{bmatrix} \in \mathbb{R}^{n\times 2} \tag{2}$$

$$pt_{3d,real} = \begin{bmatrix} X_1 & Y_1 & Z_1 \\ X_2 & Y_2 & Z_2 \\ \dots & & \\ X_n & Y_n & Z_n \end{bmatrix} \in \mathbb{R}^{n\times 3} \tag{3}$$

We can match these point sets by using unique nonlinear transformation matrix given that the surface of the gel has the hemispherical shape.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} dsin\alpha\, sin\theta \\ dsin\alpha\, cos\theta \\ dcos\alpha \end{bmatrix} \tag{4}$$

Where d is the radius of the sensor. Since we know the corresponding 3D position for the locus of observed ellipse, we can parameterize the $\alpha, \beta$ by following:

$$\alpha = a_0 + a_1 r + a_2 r^2 + ... + a_k r^k, \tag{5}$$

$$\theta = \beta\theta_0 = (b_0 + b_1 r + b_2 r^2 + ... + b_k r^k)\theta_0 \tag{6}$$

Where

$$r = \sqrt{(u-u_c)^2 + (v-v_c)^2}, \tag{7}$$

$$\theta_0 = tan^{-1}(\frac{v-v_c}{u-u_c}) \tag{8}$$

Where $u_c, v_c$ is center of the image, $(u, v)$ is corresponding pixel value from the image, with appropriate order k.

Then from the least square method, we can get the coefficient $a_i, b_i, \quad i = 0, ...k$ from solving the following:

$$\begin{bmatrix} 1 & r_1 & ... & r_1^k & 0 & ,,, & & 0 \\ 1 & r_2 & ... & r_2^k & 0 & ,,, & & 0 \\ & & ... & & & ,,, & & \\ 1 & r_N & ... & r_N^k & 0 & ,,, & & 0 \\ 0 & ,,, & & 0 & 1 & r_1 & ... & r_1^k \\ 0 & ,,, & & 0 & 1 & r_2 & ... & r_2^k \\ & & ... & & & ,,, & & \\ 0 & ,,, & & 0 & 1 & r_N & ... & r_N^k \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ ... \\ a_k \\ b_0 \\ ... \\ b_k \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ ... \\ \alpha_N \\ \beta_1 \\ ... \\ \beta_N \end{bmatrix} \tag{9}$$

From above matrix, we can get the appropriate coefficient for $\alpha, \beta$. By using above, I could get the coefficient with $k = 5, N = 300$ :

$$\begin{bmatrix} a_0 \\ a_1 \\ ... \\ a_5 \\ b_0 \\ ... \\ b_5 \end{bmatrix} = \begin{bmatrix} -0.5495 \\ 10.8267 \\ -42.4161 \\ 73.0804 \\ -49.9187 \\ 10.0703 \\ 2.1714 \\ -13.3646 \\ -8.2026 \\ 92.3806 \\ -125.4496 \\ 51.3604 \end{bmatrix}$$

From above coeffiecient with estimated parameters for the transformation, we could get the Fig. 11 as the result of 3D reconstruction. We can notice that the 1st quadrant is accurate since we only accurately matched these parameters. We can interprete the other points from the dummy pixel, since we only have to convert the pixel inside the circular lens part, but the plot draws all the pixel from 0 to maximum length of the image.

## 6. Conclusion

This paper proposed three different approach to reconstruct the 3D surface of the tactile sensor with fisheye lens from single image. The first approach, the end-to-end approch from the 3D pose of the object, shows the huge loss becaus of the small dataset( 1500). For the next approach, the 3D reconstruction from the 3D pointcloud, I was successfully calibrate the pointcloud from the three different ToF sensor by using ICP algorithm. In the future, I will train the network between image and 3D pointclouds. Finally, I proposed the 3D reconstruction from the 3D-printed indicator. By using the position of the ellipse given by Hough transform, I could match the locus of the ellipse/hole on the image pixel with the 3D position from the surface of the sensor. Finally, the 3D reconstruction has been performed with some noise. In the future, I will reduce the noise for the final approach by getting right coefficient for the nonlinear conversion matrix of the image-3D position.

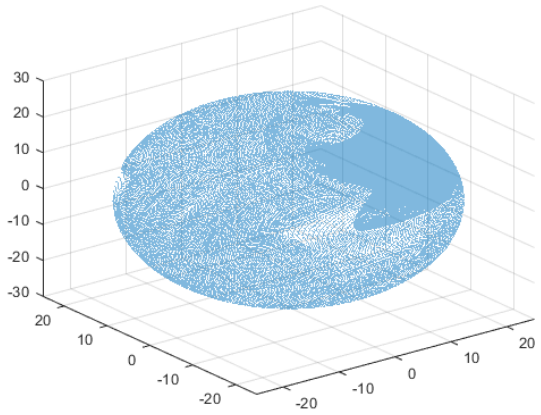


Figure 11. The 3D reconstruction by using least square method.

## References


[1] D. H. Ballard. Generalizing the hough transform to detect arbitrary shapes. *Pattern recognition*, 13(2):111–122, 1981.

[2] R. Basri and D. W. Jacobs. Lambertian reflectance and linear subspaces. *IEEE transactions on pattern analysis and machine intelligence*, 25(2):218–233, 2003.

[3] P. J. Besl and N. D. McKay. Method for registration of 3-d shapes. In *Sensor fusion IV: control paradigms and data structures*, volume 1611, pages 586–606. International Society for Optics and Photonics, 1992.

[4] J. Canny. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, (6):679–698, 1986.

[5] M. Cheng, X. Huang, C. Ma, and Y. Yang. A flexible capacitive tactile sensing array with floating electrodes. *Journal of Micromechanics and Microengineering*, 19(11):115001, 2009.

[6] S. Dong, W. Yuan, and E. H. Adelson. Improved GelSight tactile sensor for measuring geometry and slip. In *IEEE International Conference on Intelligent Robots and Systems*, volume 2017-Septe, pages 137–144. IEEE, sep 2017.

[7] E. Donlon, S. Dong, M. Liu, J. Li, E. Adelson, and A. Rodriguez. Gel Slim : A High-Resolution , Compact , Robust , and Calibrated Tactile-sensing Finger. pages 1927–1934, 2018.

[8] F. R. Hogan, M. Bauza, O. Canal, E. Donlon, and A. Rodriguez. Tactile Regrasp: Grasp Adjustments via Simulated Tactile Transformations. *IEEE International Conference on Intelligent Robots and Systems*, pages 2963–2970, 2018.

[9] T. M. Huh, H. Choi, S. Willcox, S. Moon, and M. R. Cutkosky. Dynamically reconfigurable tactile sensor for robotic manipulation. *IEEE Robotics and Automation Letters*, 5(2):2562–2569, 2020.

[10] M. Lambeta, P. W. Chou, S. Tian, B. Yang, B. Maloon, V. R. Most, D. Stroud, R. Santos, A. Byagowi, G. Kammerer, D. Jayaraman, and R. Calandra. DIGIT: A Novel Design for a Low-Cost Compact High-Resolution Tactile Sensor with Application to In-Hand Manipulation. *IEEE Robotics and Automation Letters*, 5(3):3838–3845, 2020.

[11] D. Ma, E. Donlon, S. Dong, and A. Rodriguez. Dense Tactile Force Estimation using GelSlim and inverse FEM. pages 5418–5424, 2019.

[12] A. Padmanabha, F. Ebert, S. Tian, R. Calandra, C. Finn, and S. Levine. OmniTact: A Multi-Directional High Resolution Touch Sensor. 2020.

[13] S. Pasinetti, M. M. Hassan, J. Eberhardt, M. Lancini, F. Docchio, and G. Sansoni. Performance analysis of the pmd camboard picoflexx time-of-flight camera for markerless motion capture applications. *IEEE Transactions on Instrumentation and Measurement*, 68(11):4456–4471, 2019.

[14] B. Romero, F. Veiga, and E. Adelson. Soft, Round, High Resolution Tactile Fingertip Sensors for Dexterous Robotic Manipulation. pages 4796–4802, 2020.

[15] A. Saxena, M. Sun, and A. Y. Ng. Make3d: Learning 3d scene structure from a single still image. *IEEE transactions on pattern analysis and machine intelligence*, 31(5):824–840, 2008.

[16] D. Scaramuzza, A. Martinelli, and R. Siegwart. A toolbox for easily calibrating omnidirectional cameras. In *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 5695–5701. IEEE, 2006.

[17] Y. She, S. Wang, S. Dong, N. Sunil, A. Rodriguez, and E. Adelson. Cable Manipulation with a Tactile-Reactive Gripper. 2019.

[18] N. Wettels, J. A. Fishel, Z. Su, C. H. Lin, G. E. Loeb, and L. SynTouch. Multi-modal synergistic tactile sensing. In *Tactile sensing in humanoids—Tactile sensors and beyond workshop, 9th IEEE-RAS international conference on humanoid robots*, 2009.