An Optimized Depth Complementation of Transparent Objects Based Robotic Arm Grasping System

Zhaojian Gu[1,2], Hongbo Chen[3], Ping Zhu[3], Mingyu Gao[1,2(✉)], and Yan Huang[4]

[1] School of Electronics and Information, Hangzhou Dianzi University, Hangzhou 310018, China
mackgao@hdu.edu.cn
[2] Zhejiang Provincial Key Lab of Equipment Electronics, Hangzhou 310018, China
[3] Zhejiang Fangyuan Test Group Co., Ltd., Hangzhou 311222, China
[4] China Southern Power Grid Energy Development Research Institute Co., Ltd., Guangzhou 310027, China

Abstract. In this paper, we propose a method to implement a robotic arm for grasping transparent objects and apply it to the grasping of transparent test tubes. Test tubes are one of the frequently used experimental equipment in the chemical industry, and many steps in the experimental process require the use of test tubes to hold reagents. However, as a transparent object, the test tube has unique visual characteristics, which makes it difficult for general-purpose RGB-D cameras to capture its complete depth information. To solve this problem and improve the grasping quality, we propose a robotic arm grasping system using depth completion combined with point clouds. Specifically, we propose a depth learning method to complement the original depth image of transparent objects. In addition, the coordinate transformation relationship between the camera and the robotic arm is obtained by a hand-eye calibration system, while the grasping is performed based on a point cloud map generated from the complementary depth image. Experiments show that our method can significantly improve the depth complementary performance of the transparent object images and achieve accurate grasping by the robotic arm.

Keywords: Transparent Objects Detection · Depth Completion · Robot Grasping

1 Introduction

In order for a robot to accurately perform its tasks, traditional robot programming requires strict designation of start and stop positions, which makes the robot extremely demanding for the working environment. With the rapid development of computer vision, robots are no longer limited to fixed and repetitive tasks, but can achieve automatic identification and grasping of different target objects by processing the images acquired by cameras [1, 2].

Although there are many methods to achieve object recognition [3–6], the unique visual characteristics such as reflectivity and refraction of transparent objects [7] lead

H. Yang et al. (Eds.): SmartGIFT 2022, LNICST 483, pp. 229–240, 2023.
https://doi.org/10.1007/978-3-031-31733-0_20

to the fact that most algorithms are still ineffective in recognizing transparent objects. Recent work based on deep learning has achieved good results in depth completion [8, 9], by introducing features such as surface normal and occlusion boundaries to complete the original depth [10]. However, most of the depth-completion methods ignore transparent objects and predict their true depth only from the approximation of the surface or distortion behind the transparent region, which cannot recognize the transparent object, and the lack of depth information can cause serious errors in the robot grasping system [11]. Sajjan et al. [7] optimize the initial depth estimation for the transparent object surface by introducing a mask of the transparent object, and use parallel jaws to achieve 72% grasping success rate. However, this cannot satisfy the accuracy and stability required for industrial application scenarios.

In this paper, we improve the depth complementation model for transparent objects to generate a more complete depth map, and combine the point cloud data generated from the complete depth map to obtain the position of the transparent objects in the world coordinate system [12, 13], and apply it to the robotic arm grasping system.

The main contributions of this paper are as follows:

(1) The clear grasp algorithm is improved to optimize the depth map of transparent objects to obtain the complete point cloud data, and applied to the robotic arm operating system.
(2) By processing the test tube point cloud data and combining the cubic non-uniform B-sample interpolation, the path of the robotic arm in Cartesian space is planned to achieve stable grasping of the transparent test tube. Our grasping system can achieve nearly 93% grasping success rate for transparent test tubes.

2 Related Work

2.1 Detecting Transparent Objects

The recognition of transparent objects has been a challenge in computer vision and physical fields, and traditional detection methods rely on the unique optical properties possessed by transparent objects, such as the use of charge-coupled devices [14], ground-separated stereo techniques [15], and visual shell refinement [16]. Recent methods are to predict the boundary range of transparent objects by deep learning models for their accurate localization, such as SSD models [17] and RCNN models [18]. Agastya et al. [19] proposed a transparent object segmentation method combining multimodal images and deep learning by using polarization cameras to capture multimodal maps. Sajjan et al. [7] proposed a 3D shape of transparent objects estimation method by predicting surface normal, masks and occlusion boundaries, using Cholesky optimization to estimate the 3D geometry of transparent objects in a single RGB-D image. However, the depth maps produced by these methods may not meet the stable grasping requirements of a real robot arm operating system for transparent objects. Tang et al. [20] proposed to use Generative Adversarial Networks to complement the depth maps of transparent objects, however, it requires retraining all surface normal, occlusion boundaries, and masks, which is time-consuming.

2.2 Depth Completion

With the development of depth sensing technology, commercial RGB-D cameras can achieve accurate estimation of depth, but when the surface of the object is too smooth and influenced by light, the camera obtains a depth map with a part of the data missing. Depth complementation aims to predict the missing depth data to generate a complete depth image. Zhang et al. [10] obtained geometric information of the target by predicting surface normal and occluded boundaries of RGB images, which is still insufficient for predicting depth images of transparent objects. Sajjan et al. [7] removed all depth pixels from the surface of transparent objects in the original depth map and used a linear system based on geometric constraints for depth prediction, which still had a portion of depth missing in the original depth map except for transparent objects due to the illumination environment, which was not conducive to depth reconstruction. We further processed the original depth map with a fast depth complement module (FDC) [21] to improve the accuracy of the depth complement of transparent objects.

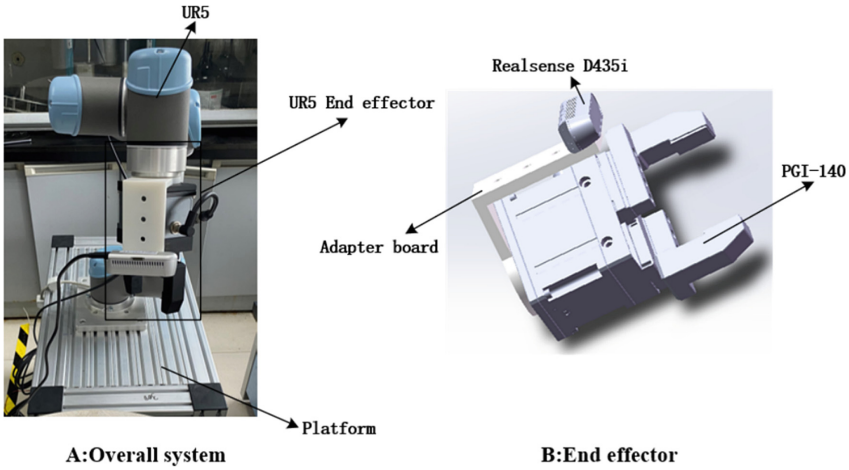

Fig. 1. Part A is the overall hardware of the robot system and part B is the end effector model of robot.

2.3 Robot Grasping

Point clouds have been applied in several ways for motion trajectory planning. Krusi et al. [22] achieved autonomous navigation in complex 3D terrain by means of unordered point cloud maps. Kuntz et al. [23], used point clouds to represent anatomical structures inside the patient's body, enabling the robot to automatically reach the surgical target inside the body while avoiding obstacles. We propose the use of point clouds to locate specific location information of transparent test tubes, and the grasping system mainly considers the trajectory based on the motion of the end gripper jaws, including the inverse kinematic solution of the robot arm based on Jacobi matrix, and a Fourier series based approximation of the robot's kinematic.

The waypoints of the grasping system are generated by the sampling-based Rapid Exploration Random Tree (RRT) algorithm [24], while the smooth trajectory is optimized by the cubic non-uniform B spline interpolation and TOPP algorithms [25]. Our system is shown in Fig. 1.A. The end-effector used by the robot arm to implement the grasping is a PGI motorized gripper, whose model is shown in Fig. 1.B.

3 Method

3.1 Depth Refinement Model

As shown in Fig. 2, given a single RGB-D image of a transparent object, surface normal, occlusion boundaries, and masks of transparent objects are first obtained from the color image. Then, the original depth image is quickly depth-complemented by a classical image processing algorithm, and the processed multiple features are used as input for global optimization to further refine the globally optimized output depth estimate.

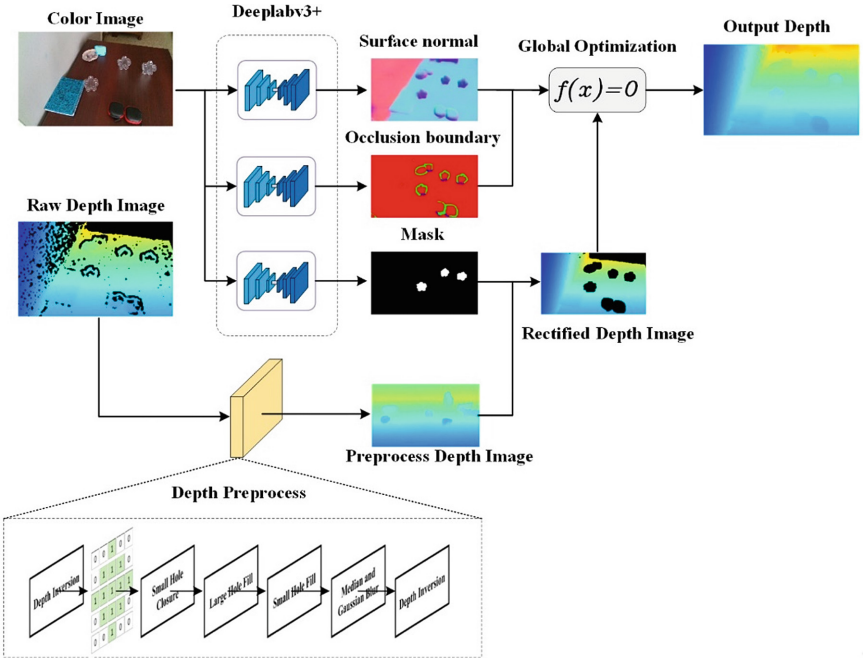

Fig. 2. Overview of our model.

Data Preprocessing. Depth complementation of transparent objects requires the use of surface normal and occlusion boundaries to provide geometric information about the object. Where the ground truth of the surface normal of a transparent object is estimated by obtaining point cloud data of an opaque object that has exactly the same shape as the transparent object and is located at the same position, while the ground truth of the occlusion boundary is predicted by the depth and mask of the same opaque object.

Initial Depth Complement. For the visual properties of transparent objects, Sajjan et al. [7] modified the original depth map by removing the masks of transparent objects. Considering the efficiency and accuracy, we first perform a fast preprocessing of the original depth map by using the classical image processing method [21] to repair the missing depth in it due to factors such as lighting environment except for transparent objects.

The process of the original depth map processing can be described as finding a function f' to approximate the true function:

$$f(I, D_{raw}) = D_{dense} \tag{1}$$

where $I \in R^{M \times N}$ is the original color image, $D_{raw} \in R^{M \times N}$ is the original depth map, and $D_{dense} \in R^{M \times N}$ is the processed depth map, which is the same size as I and D_{raw}. The process can be formulated as:

$$\min \left|\left| f'(I, D_{raw}) - f(I, D_{raw}) \right|\right|_F^2 = 0 \tag{2}$$

The function f' can be implemented by the image processing operation as shown in Fig. 3.

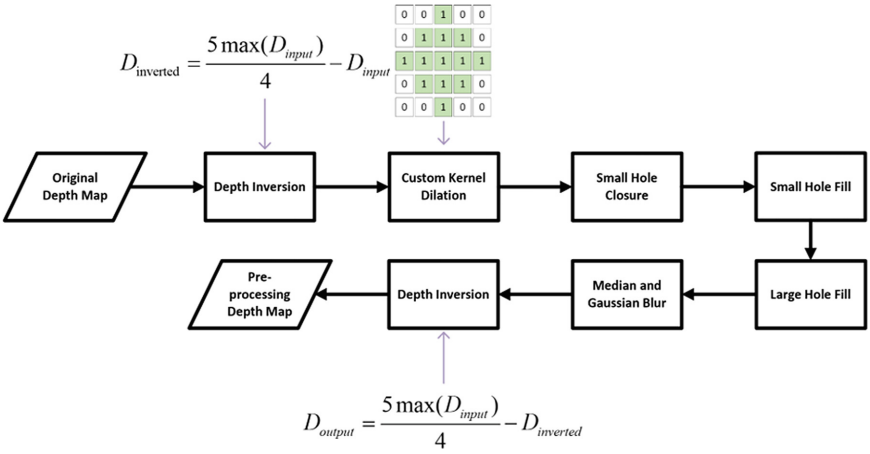

Fig. 3. Preprocess of depth map.

where D_{intered} denotes the inverted depth map, and $\max(D_{input})$ denotes taking the maximum value of the depth of the input depth map. Considering the application scenario, we removed the module that extends to the top of the frame, making it more focused on the immediate target.

Point Cloud for Grasping. We use transparent test tubes as the experimental object and first obtain the position of the transparent test tube in space from the complete depth map and generate a point cloud map. Since the test tube is placed on the test tube holder, we traverse all the point cloud data to find the highest point and use to search all the

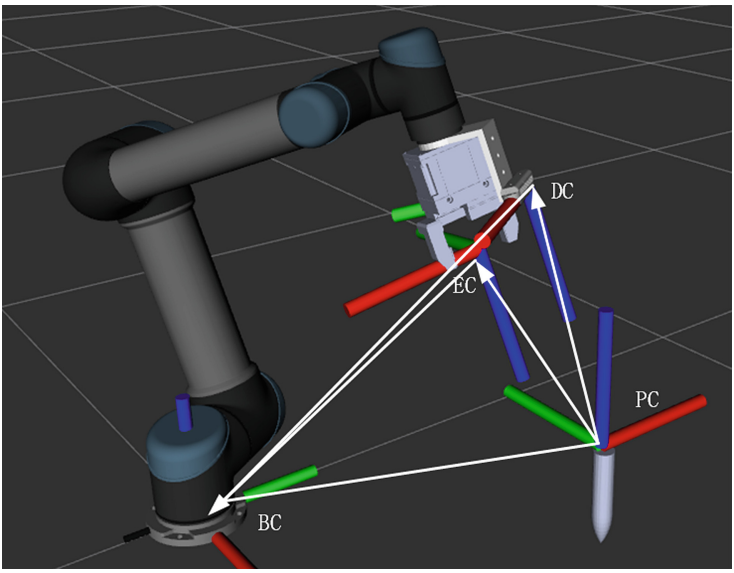

Fig. 4. Robot coordinate system.

points near the highest point that are in the same plane as the top of the test tube. Then we use the center point (x,y,z) of that plane as the grasping position of the end-effector of the robot arm.

To express our grasping method, we assume that the coordinate system of the base of the robot arm is BC, the coordinate system of the end-effector is EC, the coordinate system of the camera is DC, the coordinate system generated at the point cloud is PC, and the distribution of the coordinate systems is shown in Fig. 4. According to the established model and the hand-eye calibration system, we can infer the spatial position information of EC and DC relative to BC, and according to the point cloud information, we can obtain the spatial conversion relationship between DC and PC, so as to obtain the spatial position of DC relative to BC.

We set the spatial position and pose of the camera relative to the base as $t_{BD}^{\vec{B}}$ and R_D^B, and the spatial position and pose of the grasping position relative to the camera as $t_{DP}^{\vec{D}}$ and R_P^D, then the transformation matrix of the grasping position relative to the base coordinate system of the robot arm is as follows:

$$\begin{cases} {}^{B}T_P = {}^{B}T_E{}^{E}T_D{}^{D}T_P \\ {}^{D}T_P = \begin{bmatrix} R_P^D & t_{DP}^{\vec{D}} \\ 0 & 1 \end{bmatrix} \\ {}^{B}T_E{}^{E}T_D = \begin{bmatrix} R_D^B & t_{BD}^{\vec{B}} \\ 0 & 1 \end{bmatrix} \end{cases} \tag{3}$$

4 Experiments

4.1 Experimental Environment and Dataset

The hardware and software environments for the experiments are as follows: a server with Ubuntu 20.04, GTX3070 GPU, deep learning framework Pytorch1.11.0, Python 3.8.

We use Clear Grasp's dataset for evaluation, which contains more than 50,000 synthetic RGB-D images, as well as 286 real-world images of transparent objects. Opaque objects are used to overlay transparent objects to obtain realistic depth information. The ground truth of occlusion boundaries is generated from the depth maps of opaque objects as well as masks, and point cloud data are generated from the depth maps to obtain the ground truth of surface normal for training. We use 5 known real-world objects in Clear Grasp's dataset as testing set.

4.2 Performance Metrics

For depth-completion, we use the same metric as in the previous work: the RMSE, REL, MAE and the percentage of pixels of the ground truth where the predicted depth falls within the threshold percentage error called δ [26, 27]:

$$\delta = |\text{predicted} - \text{true}|/\text{true} \tag{4}$$

where δ is 1.05, 1.10 or 1.25.

For grasp detection, we use multiple depth completion methods for grasping, and each method is repeated 100 times, and we calculate the grasping success rate:

$$rate = \frac{successful\ picks}{picking\ attempts} \tag{5}$$

as the evaluation metric.

4.3 Comparison with Other Depth Completion Algorithm

We compare the performance of Fast Depth Completion (FDC) [21], Deep Depth Completion (DDC) [10], Clear Grasp [7] and our method, and the performance of depth completion is shown in Table 1.

We can observe that the poor performance of the model using the traditional depth-completion algorithm alone, which reasons may be that it does not consider the unique visual properties possessed by transparent objects, ignoring its local details and contour boundaries. Compared to the aforementioned methods, Clear Grasp obtains a larger performance improvement due to the geometric information obtained from the RGB image prediction and the correction of the input depth map. The best performance of our proposed method is obtained due to the further processing of depth information, which eliminates a large amount of uncertain depth information in the input depth map and achieves a more accurate optimization.

The results of depth complementation are shown in Fig. 5, and it can be observed that our method obtains relatively complete depth information. As shown in the figure, the shapes and contours of transparent objects in the circles can be well.

Table 1. Depth completion results of the approaches.

	Error Metrics			Accuracy Metrics		
Model	RMSE ↓ (m)	REL ↓ (m)	MAE ↓ (m)	$\delta_{1.05}$ ↑ (%)	$\delta_{1.10}$ ↑ (%)	$\delta_{1.25}$ ↑ (%)
FDC	0.127	0.142	0.078	48.04	66.63	88.24
DDC	0.054	0.081	0.045	44.53	69.71	95.77
Clear Grasp	0.039	0.051	0.029	72.86	86.99	95.63
Ours	0.035	0.046	0.026	75.81	89.04	96.19

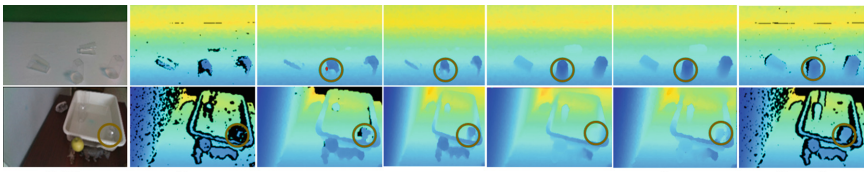

Fig. 5. Comparison of depth completion effects of multiple approaches.

4.4 Ablation Study

The preprocessing module of the depth image uses depth inversion, hole closure and fill methods to complement the depth information. We compare the effects of using different hole methods and eliminating depth inversion in the preprocessing module, and we denote the method without the depth inversion module as Ours-I, the method without small hole closure as Ours-Shc, the method without small hole fill as Ours-Shf and the method without large hole fill as Ours-Lhf. Table 2 summarizes the preprocessing performance of various variants of the module.

Table 2. Ablation results of the preprocessing module.

	Error Metrics			Accuracy Metrics		
Model	RMSE ↓ (m)	REL ↓ (m)	MAE ↓ (m)	$\delta_{1.05}$ ↑ (%)	$\delta_{1.10}$ ↑ (%)	$\delta_{1.25}$ ↑ (%)
Ours-I	0.064	0.085	0.048	49.87	74.85	92.66
Ours-Shc	0.035	0.046	0.026	75.73	88.98	96.18
Ours-Shf	0.036	0.047	0.027	75.55	88.72	95.85
Ours-Lhf	0.037	0.048	0.027	74.92	88.4	95.53
Ours	0.035	0.046	0.026	75.81	89.04	96.19

The depth map after the variant complementation by our method is shown in Fig. 6, it can be observed that the model after depth inversion obtains a clearer outline of transparent objects, while hole closure and hole fill also play a certain role.

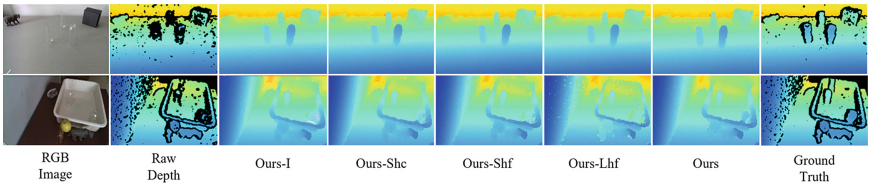

Fig. 6. Comparison to the variants of our method.

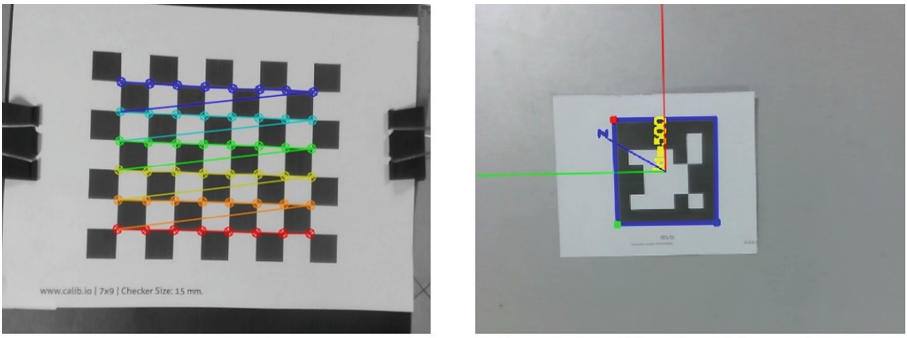

Fig. 7. Part A is camera internal parameters calibration and part B is Camera external parameters calibration.

4.5 Eye-to-Hand Calibration System

In order for the robotic arm to obtain the position of the object in the image in space, the camera needs to be calibrated. This includes the calibration of the internal parameters of the camera and the calibration of the external parameters. The internal calibration is used to correct the distortion of the image due to the distortion of the camera lens, and the external calibration is used to reconstruct the 3D scene. We use Zhang Zhengyou calibration method [28] for internal calibration, as shown in Fig. 7.A. Using a 9x7 checkerboard calibration plate of size 20 mm to calculate the internal parameters and aberration coefficients of the camera by shooting from multiple angles.

For external calibration, we use the eye-in-hand calibration method, as shown in Fig. 7.B. We move the robot arm to different positions while keeping the calibration plate within the field of view of the camera, and obtaining the motion samples of multiple points for the calculation of parameters. The specific parameters of the camera are shown in the Table 3.

4.6 Accuracy of Grasp System

We use the acquired complete depth map as input to the robot grasping system to observe the effect of depth complementation on the grasping performance of transparent objects. In this experiment, we place transparent test tubes on an experiment-specific test tube rack, capture RGB-D images by a Realsense D435i camera, and control the robotic

Table 3. Camera calibration parameters.

Internal parameters	f_x	614.367
	f_y	614.705
	x_0	325.077
	y_0	240.572
T	Position x	0.409
	Position y	0.466
	Position z	0.516
R	Orientation x	0.840
	Orientation y	−0.520
	Orientation z	0.043
	Orientation w	−0.149

arm to achieve the grasping of transparent test tubes by point cloud information. The point cloud maps obtained before and after depth complementation using our method are shown in Fig. 8, and the snapshots of the UR5 grasping transparent test tubes are shown in Fig. 9. We tested different methods to grasp the test tube 100 times, and if the error between the robot's end-effector coordinate system position and the predicted test tube coordinate position is less than 0.1 cm, the grasping is successful, and the success rate is shown in the Table 4. It can be observed that our method has the best grasping performance.

Fig. 8. Part A is the raw point cloud map and part B is the processed point cloud map

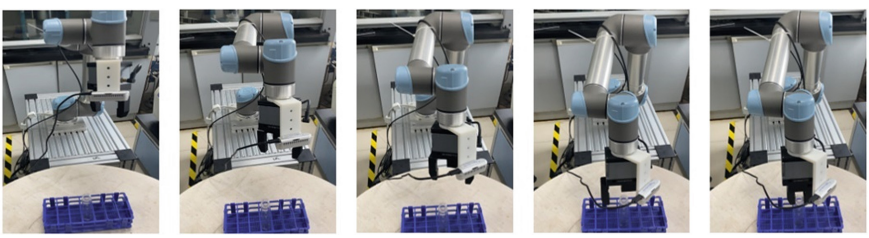

Fig. 9. UR5 grasping transparent test tube process

Table 4. Performance of grasping test tube

Method	Success rate (%)
Raw Depth	4
FDC	53
DDC	67
Clear Grasp	87
Ours	93

5 Conclusion

In this paper, we obtain more complete depth information of the transparent object by preprocessing the original depth map using the fast depth complement method, and obtain the point cloud data of the transparent object by the complete depth information to obtain its position information in the real world and control the robotic arm for grasping. The results show that the system can achieve stable grasping of transparent objects. This study is beneficial to advance the intelligence of chemical laboratory and reduce the labor cost. In future work, we will try to use the dual-arm robot to perform the tasks in chemistry experiments for test tube manipulation and to improve the robustness of the predictive occlusion boundary algorithm in complex laboratory environments.

References

1. Lu, C.: Kalman tracking algorithm of ping-pong robot based on fuzzy real-time image. J. Intell. Fuzzy Syst. **38**(4), 3585–3594 (2020)
2. Zhang, Y., Cheng, W.: Vision-based robot sorting system. In: Materials Science and Engineering, vol. 592, p. 1. IOP Publishing (2019)
3. Tanwani, A.K., Mor, N., Kubiatowicz, J.: A fog robotics approach to deep robot learning: application to object recognition and grasp planning in surface decluttering. In: International Conference on Robotics and Automation (ICRA), pp. 4559–4566. IEEE (2019)
4. Hossain, D., Capi, G., Jindai, M.: Optimizing deep learning parameters using genetic algorithm for object recognition and robot grasping. J. Electron. Sci. Technol. **16**(1), 11–15 (2018)
5. Cartucho, J., Ventura, R., Veloso, M.: Robust object recognition through symbiotic deep learning in mobile robots. In: International Conference on Intelligent Robots and Systems (IROS), pp. 2336–2341. IEEE (2018)
6. Dong, Z., Ji, X., Zhou, G.: Multimodal neuromorphic sensory-processing system with memristor circuits for smart home applications. IEEE Trans. Ind. Appl. (2022)
7. Sajjan, S., Moore, M., Pan, M.: Clear grasp: 3D shape estimation of transparent objects for manipulation. In: International Conference on Robotics and Automation (ICRA), pp. 3634–3642. IEEE (2020)
8. Sperling, L., Lämmer, S., Leipzig, S.: Uncertainty-aware evaluation of machine learning performance in binary classification tasks. J. WSCG **30**(1), 63–71 (2022)
9. Lu, K., Barnes, N., Anwar, S.: From depth what can you see? Depth completion via auxiliary image reconstruction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11306–11315. IEEE (2020)

10. Zhang, Y., Funkhouser, T.: Deep depth completion of a single RGB-D image. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 175–185. IEEE (2018)
11. Ji, X., Dong, Z., Lai, C.S.: A brain-inspired in-memory computing system for neuronal communication via memristive circuits. IEEE Commun. Mag. **60**(1), 100–106 (2022)
12. Ten Pas, A., Platt, R.: Using geometry to detect grasp poses in 3D point clouds. Rob. Res. **2**, 307–324 (2018)
13. Mahler, J., Liang, J., Niyaz, S.: Dex-Net 2.0: deep learning to plan robust grasps with synthetic point clouds and analytic grasp metrics. Robot.: Sci. Syst. (2017)
14. Jamaludin, J., Rahim, R.A., Rahiman, M.H.F.: Optical tomography system using charge-coupled device for transparent object detection. Int. J. Integr. Eng. **10**(4) (2018)
15. Phillips, C.J., Derpanis, K.G., Daniilidis, K.: A novel stereoscopic cue for figure-ground segregation of semi-transparent objects. In: International Conference on Computer Vision Workshops (ICCV Workshops), pp. 1100–1107. IEEE (2011)
16. Zuo, X., Du, C., Wang, S.: Interactive visual hull refinement for specular and transparent object surface reconstruction. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2237–2245. IEEE (2015)
17. Khaing, M.P., Masayuki, M.: Transparent object detection using convolutional neural network. In: Zin, T.T., Lin, J.-W. (eds.) ICBDL 2018. AISC, vol. 744, pp. 86–93. Springer, Singapore (2019). https://doi.org/10.1007/978-981-13-0869-7_10
18. Lai, P.J., Fuh, C.S.: Transparent object detection using regions with convolutional neural network. In: IPPR Conference on Computer Vision, Graphics, and Image Processing, p. 2 (2015)
19. Kalra, A., Taamazyan, V., Rao, S.K.: Deep polarization cues for transparent object segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8602–8611 (2020)
20. Tang, Y., Chen, J., Yang, Z.: DepthGrasp: depth completion of transparent objects using self-attentive adversarial network with spectral residual for grasping. In: 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 5710–5716. IEEE (2021)
21. Ku, J., Harakeh, A., Waslander, S.L.: In defense of classical image processing: fast depth completion on the CPU. In: 2018 15th Conference on Computer and Robot Vision (CRV), pp. 16–22. IEEE (2018)
22. Krüsi, P., Furgale, P., Bosse, M.: Driving on point clouds: motion planning, trajectory optimization, and terrain assessment in generic nonplanar environments. J. Field Robot. **34**(5), 940–984 (2017)
23. Kuntz, A., Fu, M., Alterovitz, R.: Planning high-quality motions for concentric tube robots in point clouds via parallel sampling and optimization. In: 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 2205–2212. IEEE (2019)
24. Véras, L.G.D.O., Medeiros, F.L.L., Guimaráes, L.N.F.: Systematic literature review of sampling process in rapidly-exploring random trees. IEEE Access **7**, 50933–50953 (2019)
25. Pham, Q.C.: A general, fast, and robust implementation of the time-optimal path parameterization algorithm. IEEE Trans. Robot. **30**(6), 1533–1540 (2014)
26. Huang, Y.K., Wu, T.H., Liu, Y.C.: Indoor depth completion with boundary consistency and self-attention. In: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. IEEE (2019)
27. Dong, Z., Qi, D., He, Y.: Easily cascaded memristor-CMOS hybrid circuit for high-efficiency Boolean logic implementation. Int. J. Bifurc. Chaos **28**(12), 1850149 (2018)
28. Zhang, Z.: A flexible new technique for camera calibration. IEEE Trans. Pattern Anal. Mach. Intell. **22**(11), 1330–1334 (2000)