In-Hand Object Localization Using GelStereo Visuotactile Sensing

Shaowei Cui[1], Qiang Li[2], Yinghao Cai[1], and Shuo Wang[1]

***Abstract*— In-hand object localization has always been a critical but difficult aspect of dexterous robotic manipulation. We attempt to address this issue in this paper through the use of point cloud registration techniques. Specifically, the grasping pose is estimated by registering the high-resolution 3D contact point cloud sensed by a novel GelStereo tactile sensor with the object template point cloud. Extensive qualitative and quantitative analyses of in-hand localization and insertion experiments of small parts are performed on our robot platform. The experimental results verify the accuracy and robustness of the proposed in-hand object localization pipeline.**

I. INTRODUCTION

In-hand object localization is one of the prerequisites for performing delicate and intricate robotic manipulation. [1]. Without sufficiently precise position estimation, the robot will be incapable of performing delicate manipulation on a human-level, such as assembling small parts. The majority of currently available localization methods rely on visual perception to locate the grasped object, and considerable effort has been expended on grasping configuration and environment using computer vision technologies [2]. However, due to the non-contact propriety of these vision-based approaches, they become unreliable during actual manipulations. Many issues can affect the accuracy and robustness of the vision-based localization systems, such as grasping occlusion, the field of view limits, and light deterioration [3].

Recently, tactile perception has been proven to be a reliable solution [4]. Pfanne *et al.* present an EKF-based method to estimate the grasping state using position and torque measurements from the joints of the hand [5]. Ding *et al.* describe a particle filter-based pose estimation algorithm based on tactile sensory information in combination with haptic rendering models [6]. These filtering-based methods can only provide contact level information and are modeled for specific objects, resulting in insufficient generalization and mediocre real-time performance. Bimbo *et al.* present a strategy to represent data from a tactile array sensor and match it to an object's geometric features for in-hand object pose estimation [7]. The geometry-based matching method is more versatile, but the proposed covariance-based method has poor real-time performance.

*supported in part by the National Key Research andDevelopment Program of China under Grant 2018AAA0103003.

[1]Shaowei Cui, Yinghao Cai, and Shuo Wang are with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academyof Sciences, Beijing 100190, China cuishaowei2017@ia.ac.cn; yinghao.cai@ia.ac.cn; shuo.wang@ia.ac.cn

[2]Qiang Li is with the Neuroinformatics Group & CITEC, Bielefeld University, 33619 Bielefeld, Germany qli@techfak.uni-bielefeld.de

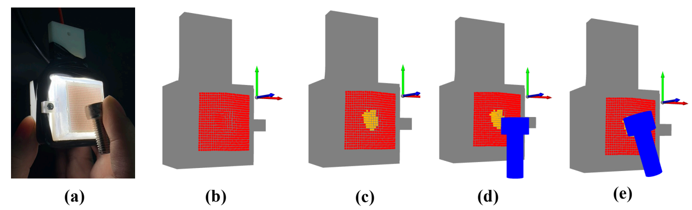

Fig. 1. The proposed in-hand object localization pipeline. (a) The GelStereo sensor contacts an M10 screw. (b) The obtained tactile point cloud. (c) The saliency point cloud detection result (yellow point cloud). (d) The initial pose of the in-hand screw. (e) The estimated pose of the contacting screw.

For contact geometry measurement, visuotactile sensors can provide ideal sensing capability [8]. GelSight sensors [9], benefit from its creative use of the photometric stereo algorithm to reconstruct depth information, which can obtain dense tactile point cloud directly. The GelSight sensor has already been applied in in-hand small parts localization and insertion task [10]. Unfortunately, this tactile images-based method needs to collect a large number of tactile images for a specific object for modeling, which is too expensive. In this paper, we propose a in-hand object pose estimation method based on 3D point cloud registration. The contact point cloud is perceived by our GelStereo tactile sensor [11], which can provide tactile contact point cloud with high spatial resolution (< 1 mm).

II. METHOD

The proposed in-hand object localization pipeline consists of two steps: saliency tactile point cloud detection and grasping pose estimation using point-set registration, as shown in Fig. 1. The saliency tactile point cloud detection process is used to extract the tactile point cloud in the contact area, and the extracted point cloud can be used to estimate the grasping pose by point-set registration.

A. Saliency Point Cloud Detection

Inspired by [12], the saliency point cloud can be extracted by setting a threshold in the contact direction, i.e. if the depth value of a tactile point on the contact surface exceeds the default threshold, this point will be considered as a saliency tactile point, the yellow dots as shown in Fig. 1(b)). Given the obtained tactile point cloud $Q \in \mathbb{R}^{N\times 3}$, the saliency point cloud $P \in \mathbb{R}^{M\times 3}$ can be obtained by:

$$\begin{aligned} &P \Leftarrow Q_i,\ if\ Q_{i,z} > z \\ &i = 1, 2, ..., N \end{aligned} \tag{1}$$

Fig. 2. Four pegs and a USB interface for grasping and insertion experiments.

where Q_i is the $i\ th$ tactile point of the tactile point cloud Q, and $Q_{i,z}$ indicates its depth value in the GelStereo coordinate system. $\Leftarrow$ denotes putting current point Q_i into the saliency point cloud P space. In practice, d is set as 0.35mm, which is determined experimentally.

B. Pose Estimation Using Point-Set Registration

After perceiving the saliency point cloud, the in-hand object's pose can be calculated by points-set registration between the saliency point-set (source) and the template point-set (target). The target object point cloud can be obtained by simple 3D modeling or using a handheld scanner device (such as iPhone 12 Pro). In this paper, a probabilistic point-set registration method named FilterReg [13] is employed.

Given a predicted (initial) in-hand pose ${}^{gs}T_{\text{init}}$, The target point-set P' is first transformed to this initial pose,

$$ {}^{gs}P'_{\text{init}} = {}^{gs}T_{\text{init}}P' \tag{2}$$

where ${}^{gs}P'_{init}$ indicate target point-set with initial pose in the GelStereo frame. Then the transformation matrix $T_r \in \mathbb{R}^{4\times4}$ of the registration between the source point-set $P \in \mathbb{R}^{M\times3}$ and the target point-set ${}^{gs}P'_{\text{init}} \in \mathbb{R}^{M'\times3}$ is computed by

$$T_r = \mathbb{F}_{\text{probreg}}(P, {}^{gs}P'_{init})\ \in \mathbb{R}^6 \tag{3}$$

where $\mathbb{F}_{\text{probreg}}(\cdot)$ indicates the FilterReg registration method, and M' is the number of points in the target point-set P'. The target point-set is then transformed to the GelStereo surface.

$$ {}^{gs}P'_r = T_r^{-1\,gs}P'_{\text{init}} \tag{4}$$

where T_r^{-1} is the inverse transformation matrix of the obtained transformation matrix T_r.

Finally, the pose of the in-hand object in the robot base frame can be transformed by:

$$ {}^{b}P'_r = {}^{b}T_g\,{}^{g}T_{gs}\,{}^{gs}P'_r \tag{5}$$

where ${}^{g}T_{gs}$ is the transformation matrix between the GelStereo frame to the gripper frame, which can be calculated by the opening width of gripper and fixed configuration of the GelStereo sensor installation.${}^{b}T_g$ describes the desired transformation matrix between the gripper frame and the robot base frame.

TABLE I
INSERTION SUCCESS RATE OF DIFFERENT PEGS

Pegs	Clearance	RANSAC+ICP	FilterReg
Peg-(a)	0.30 mm	33/36 (91.67%)	**34/36 (94.44%)**
Peg-(b)	0.25 mm	29/36 (80.56%)	**32/36 (88.89%)**
Peg-(c)	0.25 mm	27/36 (75%)	**31/36 (86.11%)**
Peg-(d)	0.20 mm	26/36 (72.22%)	**29/36 (80.56%)**
Peg-(e)	0.00 mm	2/10 (20%)	**6/10 (60%)**
All	$\leq$0.30 mm	117/154 (75.97%)	**132/154 (85.71%)**

III. EXPERIMENTS

A. Design and Setup

We equip the GelStereo sensor to our robot setup and perform actual robot experiments of random grasping and inserting different pegs. Instead of fixing the peg at the end-effector, we randomly place the target in the middle of the gripper, and the alignment process is completed by the proposed in-hand object localization method. We select four pegs and a USB interface for this peg-in-hole assembly task, as shown in Fig. 2. Peg-(a), (b), and (d) are selected inspired by the IROS robotic grasping and manipulation competition [14]. Peg-(c) is used to verify whether the proposed in-hand localization pipeline can handle non-regular object.

We chose this insertion task because the true value of the in-hand pose is difficult to obtain, and the insertion success rate can be considered an indirectly quantitative analysis of the localization accuracy. Note that this task's assembly tolerance is less than 0.3 mm and without the aid of other model-based or learning-based assembly skills.

Furthermore, we also compare the FilterReg method with the traditional RANSAC+ICP method for point-set registration in the proposed in-hand object localization pipeline. The two methods are tested on each peg 36 times and the USB interface for ten times.

B. Results

Table I presents the assembly success rate of different pegs with the two registration method. The insertion results show that the proposed FilterReg-base pose estimation method outperforms the traditional RANSAC+ICP method for all pegs and achieves an average 85.71% insertion success rate. Especially for the USB interface insertion (Peg-(e)), the RANSAC+ICP method only succeeded twice in 10 insertion experiments, while the FilterReg-based method succeeded 6 times. These results powerfully demonstrate that the probabilistic model can achieve more accurate and generalized registration than the traditional ICP-based method for the proposed in-hand object localization.

Fig. 3 shows the robot pose Rx (GelStereo pitch angle), Ry (GelStereo roll angle), GelStereo average markers motion, and perceived tactile point cloud during a successful case of aligning and inserting the USB interface. At the moment-(a), the USB interface pose is computed by the proposed in-hand localization method, and the obtained pitch angle and roll angle are 0.33 rad and -0.04 rad, respectively.

The robot then adjusts the end-effector pose Rx and Ry accordingly and complete the pose adjustment at the moment-(b). Moment-(c) and (d) indicate the robot and GelStereo point cloud states during insertion. Finally, the USB interface is successfully inserted into the base and turn on the light, as shown in Fig. 3 -Moment-(e).

IV. CONCLUSIONS

This paper propose a registration-based in-hand object localization method using GelStereo tactile sensing. We have performed extensive in-hand object grasping and insertion experiments on our robot platform. The experimental results show that the proposed in-hand object localization method achieves an 85.71% insertion success rate on five objects with different size, geometry, and clearance, which strongly indicate the accuracy and robustness of the proposed in-hand object localization method.

In the future, we will try to address the uncertainty of the point cloud registration-based methods for in-hand object localization.

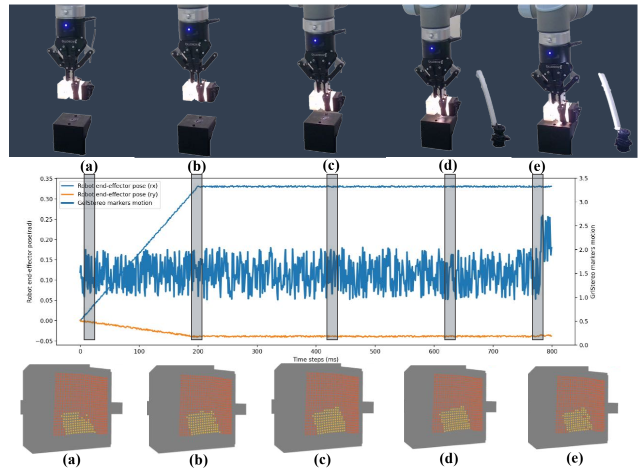

Fig. 3. The alignment and insertion process of a successful USB interface assembly experiment. Upper: Robot state sequence. Middle: Pitch angle, Roll angle, and GelStereo markers motion curve. Lower: The corresponding perceived tactile point cloud sequence.

REFERENCES

[1] A. Billard and D. Kragic, "Trends and challenges in robot manipulation," *Science*, vol. 364, no. 6446, 2019.

[2] D. Kragic, A. T. Miller, and P. K. Allen, "Real-time tracking meets online grasp planning," in *Proceedings 2001 ICRA. IEEE International Conference on Robotics and Automation (Cat. No.01CH37164)*, vol. 3, 2001, pp. 2460–2465 vol.3.

[3] C. C. Kemp, A. Edsinger, and E. Torres-Jara, "Challenges for robot manipulation in human environments [grand challenges of robotics]," *IEEE Robotics & Automation Magazine*, vol. 14, no. 1, pp. 20–29, 2007.

[4] A. S. Vazquez, R. Fernandez, A. Lopez, E. Valero, I. Payo, and A. Adan, "In-hand object localization: Simple vs. complex tactile sensors," in *SENSORS, 2014 IEEE*, 2014, pp. 1710–1713.

[5] M. Pfanne and M. Chalon, "Ekf-based in-hand object localization from joint position and torque measurements," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017, pp. 2464–2470.

[6] Y. Ding, J. Bonse, R. Andre, and U. Thomas, "In-hand grasping pose estimation using particle filters in combination with haptic rendering models," *International Journal of Humanoid Robotics*, vol. 15, no. 01, p. 1850002, 2018.

[7] J. Bimbo, S. Luo, K. Althoefer, and H. Liu, "In-hand object pose estimation using covariance-based tactile to geometry matching," *IEEE Robotics and Automation Letters*, vol. 1, no. 1, pp. 570–577, 2016.

[8] A. C. Abad and A. Ranasinghe, "Visuotactile sensors with emphasis on gelsight sensor: A review," *IEEE Sensors Journal*, vol. 20, no. 14, pp. 7628–7638, 2020.

[9] W. Yuan, S. Dong, and E. H. Adelson, "Gelsight: High-resolution robot tactile sensors for estimating geometry and force," *Sensors*, vol. 17, no. 12, p. 2762, 2017.

[10] R. Li, R. Platt, W. Yuan, A. ten Pas, N. Roscup, M. A. Srinivasan, and E. Adelson, "Localization and manipulation of small parts using gelsight tactile sensing," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2014, pp. 3988–3993.

[11] S. Cui, R. Wang, J. Hu, J. Wei, S. Wang, and Z. Lou, "In-hand object localization using a novel high-resolution visuotactile sensor," *IEEE Transactions on Industrial Electronics*, pp. 1–1, 2021.

[12] Y. She, S. Wang, S. Dong, N. Sunil, A. Rodriguez, and E. Adelson, "Cable manipulation with a tactile-reactive gripper," *arXiv preprint arXiv:1910.02860*, 2019.

[13] W. Gao and R. Tedrake, "Filterreg: Robust and efficient probabilistic point-set registration using gaussian filter and twist parameterization," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 11 095–11 104.

[14] Y. Sun and J. Falco, *Robotic grasping and manipulation: first robotic grasping and manipulation challenge, RGMC 2016, Held in Conjunction with IROS 2016, Daejeon, South Korea, October 10–12, 2016, Revised Papers*. Springer, 2018, vol. 816.