Using collocated vision to improve tactile sensing

Arkadeep Narayan Chaudhury
Robotics Institute
Carnegie Mellon University
arkadeepnc@cmu.edu

Timothy Man
Dept. of Mechanical Engineering
Carnegie Mellon University
tman2@andrew.cmu.edu

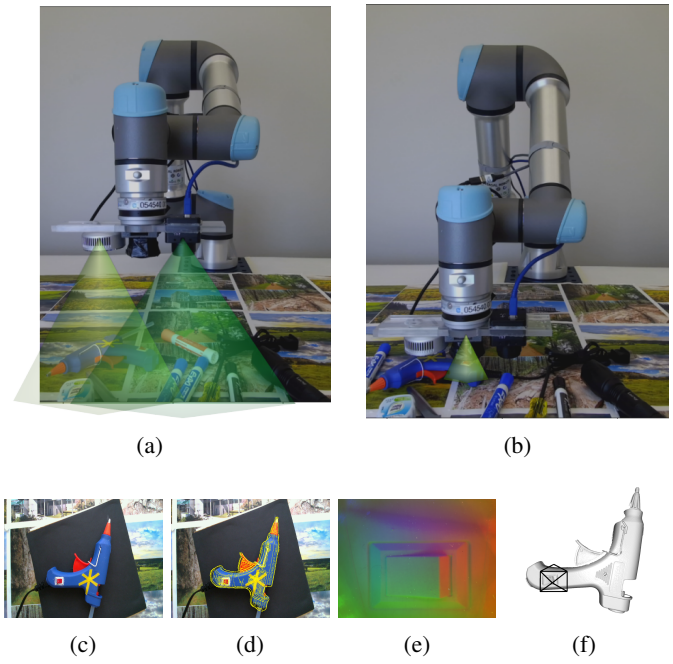
Wenzhen Yuan
Robotics Institute
Carnegie Mellon University
yuanwz@cmu.edu

Christopher Atkeson
Robotics Institute
Carnegie Mellon University
cga@cmu.edu

***Abstract*—A key step in manipulation is estimating the points of contact and pose of an object with respect to the robot. In this work we make the point that coordinating tactile sensing with vision by collocated cameras (instead of only considering the tactile sensor inputs alone) can 1) provide useful information in advance of contact and 2) simplify the contact point and pose estimation problem. We divide the problem of contact pose estimation into two parts – the initial phase where the end effector is not in contact with the object and the final phase when the end effector is in contact. We leverage the natural scopes of the camera and the tactile sensor to visually servo towards the object and obtain a coarse pose estimate of the object with respect to the end effector and refine that estimate to localize the contact using the data obtained from the tactile sensor.**

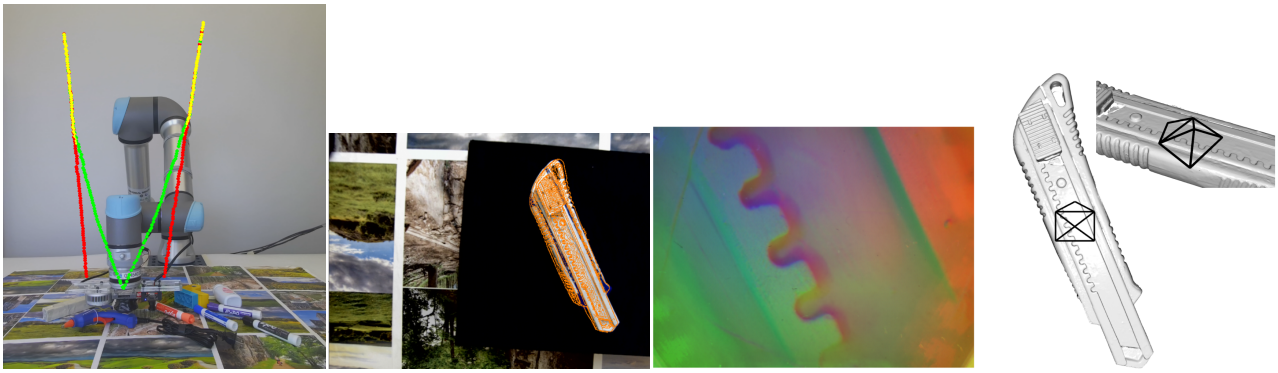
I. Introduction

In prior work on manipulation, we have found that head-mounted or external cameras are often occluded by the robot or other objects, and that objects move due to contact and during graspings so these cameras alone often cannot accurately predict contact location. Cameras that do not move with the robot hand do not get the benefit of a) direct measurement of object locations and poses relative to the hand, b) direct measurement of hand motion relative to the (potentially unknown or moving) object, and c) direct measurement of the hand location relative to the approach axis, which is useful for centering the hand with respect to the object and guiding the hand to a particular contact location. External cameras need to use stereo, multi-view, or other forms of depth measurement to locate the hand relative to the approach axis. Recent work on addressing these issues have used hand mounted cameras to demonstrate superior performance in classical manipulation tasks such as grasping and bin picking (see e.g. Song et al. [1]). With the availability of a visual perspective complementary to external (or head mounted) cameras, researchers have diversified the moving cameras to serve as tactile devices (see e.g. Yamaguchi and Atkeson [2], Yuan et al. [3]) and have implemented delicate manipulation behaviors (see e.g. Yamaguchi and Atkeson [4], Yuan et al. [5]). Recent research has also implemented tactile sensors for estimating contact pose and inferring objects from contacts (see e.g. Wang et al. [6], Smith et al. [7]), tracking object motion by fusing externally mounted cameras and tactile sensors (see e.g. Izatt et al. [8]), on transfer of information between external cameras and hand-mounted cameras (see e.g. Li et al. [9]) and for surface crack detection (see e.g. Palermo et al. [10]). A closely related work by Luo et al. [11] discuss integration of a visual and tactile measurements through a recursive Bayesian filter.



(a) (b) (c) (d) (e) (f)

Fig. 1: We demonstrate a pipeline to integrate two vision based sensors with different fields of view to visually servo the robot arm to a predetermined contact point and estimate the pose of a fixed object relative to the sensors at contact. Figures 1a and 1b show our sensor platform. We use 2 cameras – a LiDAR based RGBD sensor (Intel RealSense L515) with a 70° field of view to provide depth (fig. 1a left), and a USB camera (ELP camera with a Sony IMX 291 sensor) with a wider 100° field of view lens (fig. 1a right), which we collocate with a camera based tactile sensor – GelSight in the middle (fig. 1b).In our modified version of the GelSight, where we implement the working of the original GelSight with 6 independently controlled lights as described by Johnson et al. [12] in the physical sensor form factor introduced by Yuan et al. [3]. The cameras are used to visually servo the robot and record data (fig. 1c) and generate a preliminary pose estimate (fig. 1d) while the robot is moving towards the target. At contact, the GelSight data is observed (fig. 1e) and the preliminary pose is then refined to generate the object pose at contact. Figure 1f shows the camera pose super-imposed on the mesh model of the object.



(a) (b) (c) (d)

Fig. 2: A summary of our main results. Figure 2a shows the results of correcting trajectory errors using the optical flows observed by the two collocated cameras. On the two sides of the robot, we show 2 trajectories where the robot travels about 1 m (vertically) from the start to the final contact position, and each of the trajectories need a correction of 10 cm errors in X (horizontally left to right) and Y (horizontally into the image) directions. We show the corrected portions of the trajectories in green, un-corrected portions in red, and common portions in yellow. We could reliably correct up to 5 mm in horizontal directions at the table surface. Figures 2b to 2d describe how collocated vision can help disambiguate localization when the tactile signals are ambiguous. We set up the experiment to touch the portion of the box cutter with repeated features (the middle of slider teeth in this case). Figure 2b shows a box cutter mesh registered to the image captured by the camera, fig. 2c shows the GelSight image captured at contact and fig. 2d is the pose of the tactile sensor at contact rendered on the mesh model of the object. As the camera is physically close to the tactile sensor, we could use the pose estimated in the camera frame (fig. 2b) and refine it to predict the contact geometry recorded by the GelSight in fig. 2d, thus solving the contact pose estimation problem. Localizing this contact, in the absence of the prior pose estimates based on a wider field of view, is a harder problem. We can visually verify the success of the registration by comparing the raw GelSight data and the GelSight view at contact in figs. 2c and 2d

In this work we make the point that coordinating tactile sensing with vision by a set of collocated cameras[1] of different field of views, instead of only considering the tactile sensor inputs alone can 1) provide useful information in advance of contact and 2) simplify the contact point and pose estimation problem. To do this, we divide the problem of contact pose estimation into two parts – the initial phase before contact, when the cameras can be used for vision-based servoing to a contact point target as well as estimating a prior for contact point and object pose estimation and the final phase which refines the prior pose estimates through contact. In this paper we assume that 1) the object is not moving, 2) the object is a single rigid body with no articulations, and 3) we have a prior (potentially imperfect) 3D model of the object (potentially provided by our vision of the current object) so we can express the pose of the object with respect to this model. For this paper we put aside the gross object localization and recognition problem in order to focus on fine localization, so we assume a vision system has already located the object, created a bounding box, and recognized the object by creating or selecting an appropriate 3D model that we want to register the actual object to. Our experimental pipeline involves selecting a workspace goal and then visually servoing to that goal, recording color and depth data from the vision sensors, generating and maintaining pose estimates of the object, and using the estimates along with the tactile information received at contact to localize the contact point with respect to the robot. Through this work we show that:

- The optic flow, as observed by the hand mounted cameras, can be used to predict the heading direction of the robot. These predictions, when averaged across the cameras and portions of the robot trajectories (~ 10)cm, can provide reliable estimates of trajectory error, if any, in the robot frame. Optic flow from shorter movements was not reliable in predicting contact points due to small physical rotations of the camera caused both by interpolating the robot inverse kinematic solutions along the trajectory as well as unmeasured motions of the hand relative to the wrist joint angle sensors such as gear backlash and play.
- A couple of "mid-course" corrections in the trajectory errors estimated through the optical flows measured by the 2 cameras can correct almost all the error in trajectories while trying to move to a desired contact point (see fig. 2a).
- Pose orientation errors, when measured only with the cameras about a vertical axis are less than 1° in rotation, and 1 cm in translation, at a distance of about 30 cm from the camera.
- Given these priors, tactile estimation based on a GelSight

[1]In this work we collocate vision and touch by simply putting the cameras and the camera based tactile sensor in close physical proximity, while operating them independently.

sensor further improved the pose estimates to an uncertainty of ± 1.5mm and $\pm 0.25^{\circ}$.

- Collocated vision is particularly useful when an object does not have distinctive surface texture, or has repetitive surface texture. We show that using tactile sensing collocated with vision can help disambiguate tactile signals when used for localization.

References

[1] Shuran Song, Andy Zeng, Johnny Lee, and Thomas Funkhouser. Grasping in the wild: Learning 6dof closed-loop grasping from low-cost demonstrations. *IEEE Robotics and Automation Letters*, 5(3):4978–4985, 2020.

[2] Akihiko Yamaguchi and Christopher G Atkeson. Combining finger vision and optical tactile sensing: Reducing and handling errors while cutting vegetables. In *2016 IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids)*, pages 1045–1051. IEEE, 2016.

[3] Wenzhen Yuan, Siyuan Dong, and Edward H Adelson. Gelsight: High-resolution robot tactile sensors for estimating geometry and force. *Sensors*, 17(12):2762, 2017.

[4] Akihiko Yamaguchi and Christopher G Atkeson. Implementing tactile behaviors using fingervision. In *2017 IEEE-RAS 17th International Conference on Humanoid Robotics (Humanoids)*, pages 241–248. IEEE, 2017.

[5] Wenzhen Yuan, Mandayam A Srinivasan, and Edward H Adelson. Estimating object hardness with a gelsight touch sensor. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 208–215. IEEE, 2016.

[6] Shaoxiong Wang, Jiajun Wu, Xingyuan Sun, Wenzhen Yuan, William T Freeman, Joshua B Tenenbaum, and Edward H Adelson. 3D shape perception from monocular vision, touch, and shape priors. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1606–1613. IEEE, 2018.

[7] Edward J Smith, Roberto Calandra, Adriana Romero, Georgia Gkioxari, David Meger, Jitendra Malik, and Michal Drozdzal. 3D shape reconstruction from vision and touch. *arXiv preprint arXiv:2007.03778*, 2020.

[8] Gregory Izatt, Geronimo Mirano, Edward Adelson, and Russ Tedrake. Tracking objects with point clouds from vision and touch. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4000–4007. IEEE, 2017.

[9] Yunzhu Li, Jun-Yan Zhu, Russ Tedrake, and Antonio Torralba. Connecting touch and vision via cross-modal prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10609–10618, 2019.

[10] Francesca Palermo, Jelizaveta Konstantinova, Kaspar Althoefer, Stefan Poslad, and Ildar Farkhatdinov. Implementing tactile and proximity sensing for crack detection. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 632–637. IEEE, 2020.

[11] Shan Luo, Wenxuan Mou, Kaspar Althoefer, and Hongbin Liu. Localizing the object contact through matching tactile features with visual map. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3903–3908. IEEE, 2015.

[12] Micah K Johnson, Forrester Cole, Alvin Raj, and Edward H Adelson. Microgeometry capture using an elastomeric sensor. *ACM Transactions on Graphics (TOG)*, 30(4):1–8, 2011.