Tactile-Vision Integration for Task-Compatible Fine-Part Manipulation

Mabel M. Zhang
GRASP Laboratory
University of Pennsylvania
zmen@seas.upenn.edu

Renaud Detry Larry Matthies
Jet Propulsion Laboratory
California Institute of Technology
renaud.j.detry@jpl.nasa.gov

Kostas Daniilidis
GRASP Laboratory
University of Pennsylvania
kostas@cis.upenn.edu

***Abstract*—We propose to integrate tactile and visual sensing to predict task-compatible grasp regions for manipulation. We address the problem of fine-part assembly, by leveraging vision to observe scene-level information, then touch to perceive fine local details necessary for task completion. We directly target the end goal of task success, by predicting grasp regions that are simultaneously geometrically stable and task-compatible. We represent grasp regions with 2D probabilistic maps, which we first coarsely estimate with vision, and then refine by making contacts with the scene. We show preliminary results of probabilistic grasp regions generated by vision, and demonstrate the impact of tactile sensors for disambiguating task-compatible regions.**

I. Introduction

Tactile sensing and vision are two synergetic modalities for manipulation. While vision provides us with scene-wide observations, tactile sensing allows us to gather data in areas that are occluded from the camera by the object or the gripper and would not be accessible otherwise. Leveraging tactile sensing early in a manipulation task has a dual impact: it allows the robot to verify the reliability of scene-level properties inferred from vision, and it gives the robot an opportunity to adjust the end-effector's configuration from contact information.

In this paper, we address the problem of task-compatible grasp pose planning, whereby the robot generates grasps suitable for the intended task. We leverage touch to disambiguate parameters inferred from vision, both for task compatibility and geometrical feasibility. Our approach is novel in two parts. First, unlike traditional grasp planning approaches that reason about contact mechanics using physical constraints, we infer grasp poses directly from images and tactile readings. Second, we define the goal directly in terms of task satisfaction, instead of the individual grasp itself. This work extends our previous work in vision-based grasp planning [11], task-oriented grasping [12], and tactile-based exploration [29].

Vision and touch should be integrated in such a way as to leverage the advantages of each modality. Vision has a global field of view, with mature algorithms to interpret the scene in typical indoor lighting conditions. Touch is inherently local and time-consuming and is more suitable for fine-grained low-level sensing for tasks that require accurate interactions. We propose to use this integration for fine-part assembly, where vision is first used to perceive scene-level information for a grasping task, with ambiguity in object pose due to object shape and the intended task (Fig. 1). The output from the vision system is in the form of a probabilistic heat map (henceforth referred to as "graspability mask") of candidate grasp locations that are geometrically suitable. Subsequently, the uncertainty of each pixel in the graspability mask is computed, resulting in a second mask of probabilistic variance ("uncertainty mask"), which can be used as a cue to move the end-effector to contact uncertain regions in the scene.

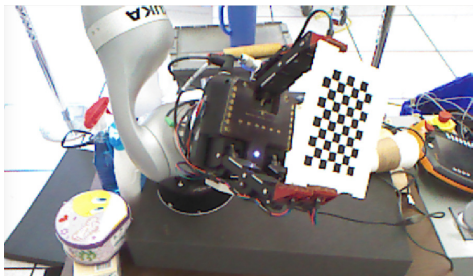

Fig. 1. Robot camera image showing ambiguity in visually perceived object pose. The object is a 3D-printed piece of a larger assembly. Depending on the stage of assembly, one of three sides of the object is not task-compatible (*i.e.* should be avoided by the gripper), as it contains extrusions and indents for attaching to another piece. From a camera, the small indents are not obvious, and the sides are indistinguishable. This image is created with the object raised to be in front of the camera; the ambiguity worsens when the object starts out on the table. Tactile sensors are able to distinguish indents.

The purpose of the tactile exploration is to improve our confidence in the grasp regions predicted by vision, in an attempt to, for instance, find regions with high potential of task success but mistakenly labeled otherwise by vision, or correct regions with low actual task success but evaluated as high by vision. During the touch sequence, the graspability mask and uncertainty mask are updated to reflect the information obtained from physically probing the scene. At the end of the sequence, a movement to actually perform the task is computed from the final belief maps.

Related work are in several categories: tactile-based manipulation, vision-based grasp planning, tactile-vision fusion, and task-oriented grasping. Pertaining to manipulation, tactile sensing has been used for grasping [2, 8, 10, 7, 18, 23, 6], recognition [25, 27, 28, 29], localization [15], to name a few

The research described in this paper was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. This research was sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-10-2-0016.

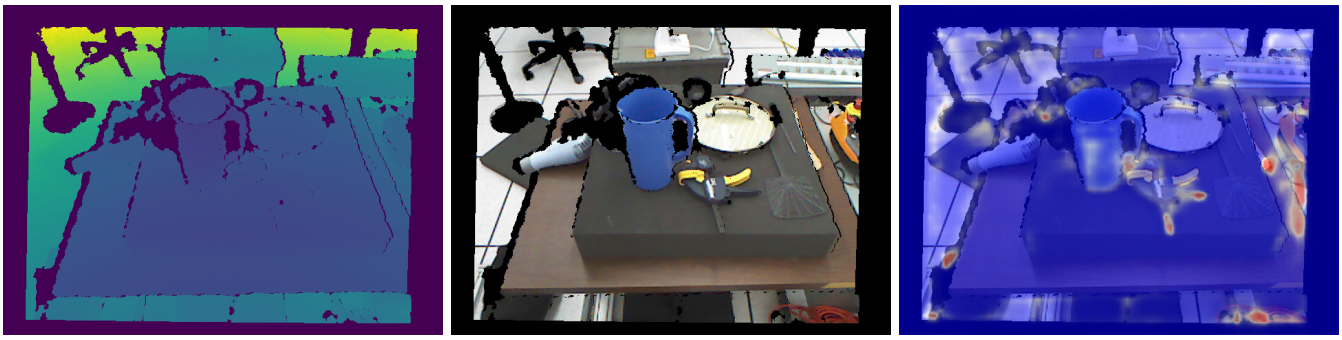

Fig. 2. Left: Depth image input to CNN, scaled to show in color. Middle: Color overlaid on point cloud depth image. Right: Probabilistic map output by CNN, highlighting regions that resemble handles for grasping.

most recent work. A survey of visual data-driven grasping is given in Bohg et al. [4]. More recently, convolutional neural networks have been used in vision-based grasping [26, 22, 21, 20]. However, the purely 2D pixel-based mapping is insufficient for manipulation beyond simple picking and poking. Tactile sensing and vision have been integrated for recognition, grasping, in-hand adaptation, and shape reconstruction [5, 14, 13, 19]. While these methods approach manipulation as a step-by-step problem, solving perception and grasping as separate subproblems, we directly target task success as the goal. In the area of task-oriented manipulation, related work include task-based quality metric [24], tactile exploration [17], vision-based grasping and adaptation [16, 1, 3], and vision-tactile grasping [9]. We combine vision and touch and reason about the uncertainty of task success in terms of pixel regions, as opposed to explicitly estimating object pose in [17]. We do not use handcrafted image features to map to grasps like [9].

II. Approach

Given the camera input of a tabletop scene, we first extract an initial 2D graspability mask of grasping regions based on geometric stability between the object and the gripper. A second mask representing the uncertainty learned from the graspability mask is generated. These maps are used to calculate a prior that determines locations that need to be explored by touch. The locations are driven by task compatibility. As the robot moves to touch the scene, the maps are updated. Finally, a task-completing movement is computed from the maps at the end of the tactile probing sequence. Detailed steps are discussed in this section.

A. Vision-Initialized and Touch-Refined Probabilistic Maps

We generate a probabilistic map of candidate grasping regions with a convolutional neural network (CNN) trained in simulation to find gripper-fitting shapes in 2D images. Fig. 2 shows an example scene and probabilistic map generated. Note the ambiguity in the coarse edges, even in large object parts. A second variance map that reflects the uncertainty of the first probabilistic map can also be outputted from the CNN. This tells us how reliable the visually detected graspable regions are. The visual procedure ends here, and the following step is to use tactile sensing to improve regions with high uncertainty.

The uncertainty mask is used as a prior to inform us where more information is needed. These are the regions we should move the end-effector to make contacts with the scene. A set of end-effector poses can be generated in the regions with high local maxima in terms of the tradeoff between graspability and uncertainty. One way is by taking $\arg\max_x g(x) + w \cdot q(x)$, where $x \in \mathbb{R}^2$ is a pixel location, $g(x)$ is the value at x in the graspability mask, $q(x)$ the value in the uncertainty mask, and w is a chosen importance weight between the two masks.

A top few local maxima can be found by thresholding. These points represent peaks of regions that have a high enough combined graspability and uncertainty to be worthy of tactile exploration. The 2D pixels at the local maxima can be mapped to corresponding 3D points in the world frame by RGBD camera parameters. Based on the geometric shape of the region surrounding the maxima, a gripper pose can be estimated by a shape matching grasp planner [11] in existing literature. Each point is explored by touch, and tactile readings are used to update the masks.

At each point chosen to be contacted, the gripper is moved to the planned pose and closed. The presence of an object inside the gripper can be detected by whether the fingers have closed all the way - finger distance to finger in the case of a pinch grasp, or finger distance to palm in the case of a wide grasp. If the enclosure is empty, then the graspability mask is updated with a low value, and the uncertainty mask with a low uncertainty. If the enclosure is non-empty, then the tactile and joint readings are recorded, a feature descriptor is constructed, and a trained discriminator (Sec. II-B) is used to determine the task compatibility of the descriptor. This task compatibility is in range $[0, 1]$ and is used to update the graspability mask, by multiplication with a Gaussian centered at x with height proportional to the task compatibility. The uncertainty mask is updated with a low uncertainty.

The tactile updates inform us about two properties of the explored regions. Directly, the discriminator trained on task compatibility tells us whether a region should be avoided for the task. Implicitly, the actual physical contact tells us whether a region will produce a geometrically stable grasp. In

the end, after the tactile updates to the graspability mask, it contains both geometric stability as well as task compatibility information from the tactile stage.

B. Touch-based Task Discriminator

To distinguish between tactile signatures that indicate whether a grasp is task-compatible, we trained a regression model on tactile descriptors with binary task compatibility labels. The tactile feature descriptor contains readings from the tactile sensor and joint angles of the gripper. The exact feature vector is hardware-dependent and specified in Sec. III.

During an assembly task, two pieces are joined by some region on each piece. These regions should be avoided by the gripper, to keep them available to be attached. We consider assembly pieces with indents, gaps, and holes, such as in Fig. 3, which are detectable by our sensing hardware. The low resolution in the hardware poses limits on the type of objects it is able to distinguish. Given finer grained sensors, objects with finer detail and variety of materials can be considered.

In the training stage, the gripper is placed in various grasping configurations with respect to the object, covering compatible and incompatible cases. Example configurations are shown in Fig. 5. The sensor inputs from each configuration are recorded as a feature descriptor and given a binary label, 0 for incompatible, 1 for compatible.

For a given task, the regression model evaluates descriptors for both task compatibility and gripper pose. The former is explicitly given in the binary training label. The latter is object-dependent and linked to which edges of the object are available to be grasped during an attachment task. The edges involved in the attachment, and therefore to be avoided by the gripper, impose implicit constraints on gripper poses. A gripper pose compatible with the task would need to place fingers only on the available edges. Such poses are defined by the gripper's geometry and kinematics, which can be solved by a geometric shape-matching grasp planner [11].

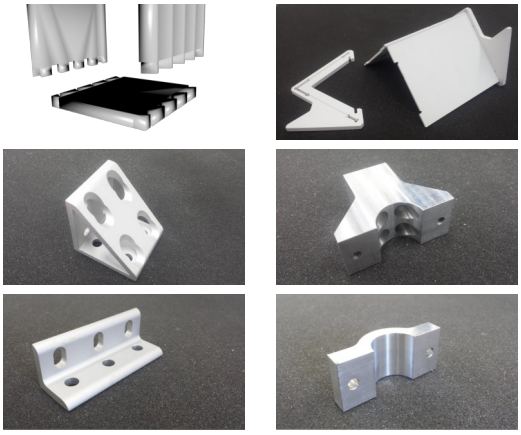

Fig. 3. Example indents, gaps, and holes commonly found in assembly parts.

III. Experiment Results

We show results of a CNN-generated probabilistic map of grasping regions (Fig. 2) and task compatibility as judged by a tactile discriminator.

The robot platform we used is shown in Fig. 1, a Robotiq 3-finger gripper equipped with 36 Takktile MEMS barometric sensors on the fingertips and palm, attached to a KUKA LBR iiwa 14 R820 arm. Each fingertip has a 2×3 array of sensors, with sensors spaced 5 mm apart. The palm has 18 sensors but rarely makes contact with small objects in a pinch grasp.

Fig. 4 shows tactile readings on an incompatible and a compatible grasp, captured on the aluminum triangle block in Fig. 3. The assembly task needs to avoid the large holes on the diagonal face. The incompatible grasp was produced by placing the fingers on the diagonal face with the large holes, and the compatible grasp placed the fingers on the sides (Fig. 5). In the incompatible grasp, the thumb shows two sensors on the object surface and one on the hollow portion. In the compatible grasp, the diagonal edges produced contiguous sensor activations.

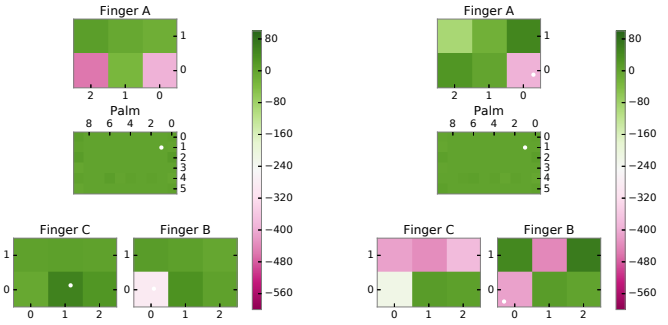

Fig. 4. Tactile readings for an incompatible (left) and a compatible (right) grasp, arranged in the shape of the gripper. Top: thumb; Center: palm; Bottom: forefingers. Medium green: no contact; dark pink: heavy contact.

The tactile feature descriptor is 46-dimensional: 36 Takktile readings and 10 gripper joint angles (4 controllable, 6 compliant). We evaluated the tactile features by recording samples at different orientations and placements of the gripper with respect to the object. For preliminary results, the samples were done by manually placing the object at different poses into the gripper. Fig. 5 shows examples of task-compatible grasps for each object. A logistic regression model is learned. The samples are randomly split into 50% training set and 50% test set, and we report the average out of 100 splits.

We recorded tactile features for the aluminum objects in Fig. 3, 25 samples on the triangle block, 24 on the joint, 9 on the 90-degree plate, and 13 on the pipe bracket. Separately for each object, the model's prediction accuracy for compatibility was 87.8%, 65.9%, 90.8%, 68.5% respectively, and overall 67.7%.

IV. Conclusion

We showed results from individual subparts of the proposed visual-tactile integration. From the graspability mask, an uncertainty mask can be learned to initialize the touch sequence. During contacts, using the tactile discriminator, the masks can be updated and used to predict the final task-completing move.

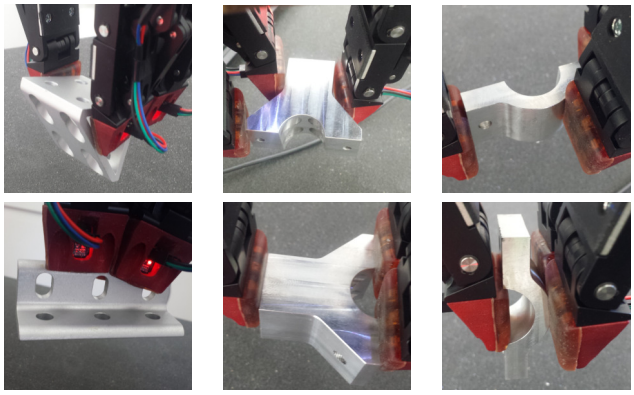

Fig. 5. Example compatible and incompatible grasps. Col 1: compatible grasps for two objects. Col 2, 3: compatible (top) and incompatible (bottom) grasps for one object per column.

REFERENCES

[1] Amir M. Ghalamzan E., N. Mavrakis, M. Kopicki, R. Stolkin, and A. Leonardis. Task-relevant grasp selection: A joint solution to planning grasps and manipulative motion trajectories. In *IROS*, 2016.

[2] Y. Bekiroglu, R. Detry, and D. Kragic. Grasp stability from vision and touch. In *Advances in Tactile Sensing and Touch-based Human Robot Interaction (IROS Workshop)*, 2012.

[3] J. Bohg, K. Welke, B. León, M. Do, D. Song, W. Wohlkinger, A. Aldoma, M. Madry, M. Przybylski, T. Asfour, H. Marti, D. Kragic, A. Morales, and M. Vincze. Task-Based Grasp Adaptation on a Humanoid Robot. In *IFAC Symposium on Robot Control*, 2012.

[4] J. Bohg, A. Morales, T. Asfour, and D. Kragic. Data-Driven Grasp Synthesis - A Survey. *TRO*, 2014.

[5] J. Bohg, K. Hausman, B. Sankaran, O. Brock, D. Kragic, S. Schaal, and G. S. Sukhatme. Interactive Perception: Leveraging Action in Perception and Perception in Action. *arXiv:1604.03670*, 2016.

[6] Y. Chebotar, K. Hausman, O. Kroemer, G. S. Sukhatme, and S. Schaal. Regrasping using Tactile Perception and Supervised Policy Learning. In *AAAI Symposium on Interactive Multi-Sensory Object Perception for Embodied Agents*, 2017.

[7] H. Dang and P. Allen. Stable Grasping under Pose Uncertainty Using Tactile Feedback. In *AURO*, 2014.

[8] H. Dang and P. K. Allen. Learning grasp stability. In *ICRA*, 2012.

[9] H. Dang and P. K. Allen. Semantic grasping: Planning robotic grasps functionally suitable for an object manipulation task. In *IROS*, 2012.

[10] H. Dang and P. K. Allen. Grasp adjustment on novel objects using tactile experience from similar local geometry. In *IROS*, 2013.

[11] R. Detry, C. H. Ek, M. Madry, and D. Kragic. Learning a Dictionary of Prototypical Grasp-predicting Parts from Grasping Experience. In *ICRA*, 2013.

[12] R. Detry, J. Papon, and L. Matthies. Semantic and geometric scene understanding for task-oriented grasping of novel objects from a single view. In *Learning and control for autonomous manipulation systems: the role of dimensionality reduction (ICRA Workshop)*, 2017.

[13] K. Hang, M. Li, J. A. Stork, Y. Bekiroglu, F. T. Pokorny, A. Billard, and D. Kragic. Hierarchical Fingertip Space: A Unified Framework for Grasp Planning and In-Hand Grasp Adaptation. *TRO*, 2016.

[14] K. Hausman, C. Corcos, J. Müller, F. Sha, and G. S. Sukhatme. Towards Interactive Object Recognition. In *3rd Workshop on Robots in Clutter: Perception and Interaction in Clutter, IROS*, 2014.

[15] P. Hebert, T. Howard, N. Hudson, J. Ma, and J. W. Burdick. The next best touch for model-based localization. In *ICRA*, 2013.

[16] M. Hjelm, R. Detry, C. H. Ek, and D. Kragic. Representations for cross-task, cross-object grasp transfer. In *ICRA*, 2014.

[17] K. Hsiao, L. P. Kaelbling, and T. Lozano-Péerez. Task-driven tactile exploration. In *RSS*, 2010.

[18] E. Hyttinen, D. Kragic, and R. Detry. Learning the Tactile Signatures of Prototypical Object Parts for Robust Part-based Grasping of Novel Objects. In *ICRA*, 2015.

[19] J. Ilonen, J. Bohg, and V. Kyrki. Fusing visual and tactile sensing for 3-D object reconstruction while grasping. In *ICRA*, 2013.

[20] E. Johns, S. Leutenegger, and A. J. Davison. Deep Learning a Grasp Function for Grasping under Gripper Pose Uncertainty. In *IROS*, 2016.

[21] D. Kappler, J. Bohg, and S. Schaal. Leveraging Big Data for Grasp Planning. In *ICRA*, 2015.

[22] I. Lenz, H. Lee, and A. Saxena. Deep learning for detecting robotic grasps. *IJRR*, 2015.

[23] M. Li, Y. Bekiroglu, D. Kragic, and A. Billard. Learning of grasp adaptation through experience and tactile sensing. In *IROS*, 2014.

[24] Y. Li, J. L. Fu, and N. S. Pollard. Data-driven grasp synthesis using shape matching and task-based pruning. *IEEE Transactions on Visualization and Computer Graphics*, 2007.

[25] Z. Pezzementi, E. Plaku, C. Reyda, and G.D. Hager. Tactile-Object Recognition From Appearance Information. *TRO*, 2011.

[26] L. Pinto and A. Gupta. Supersizing self-supervision: Learning to grasp from 50K tries and 700 robot hours. In *ICRA*, 2016.

[27] A. Schneider, J. Sturm, C. Stachniss, M. Reisert, H. Burkhardt, and W. Burgard. Object identification with tactile sensors using bag-of-features. In *IROS*, 2009.

[28] M. M. Zhang, M. Kennedy, M. Ani Hsieh, and K. Daniilidis. A Triangle Histogram for Object Classification by Tactile Sensing. In *IROS*, 2016.

[29] M. M. Zhang, N. Atanasov, and K. Daniilidis. Active Tactile Object Recognition by Monte Carlo Tree Search. *arXiv:1703.00095*, 2017.