Using Vision for Pre- and Post-Grasping Object Localization for Soft Hands

Changhyun Choi, Joseph DelPreto, and Daniela Rus

Computer Science & Artificial Intelligence Lab
Massachusetts Institute of Technology
Cambridge, MA 02139, USA
{cchoi,delpreto,rus}@csail.mit.edu
http://people.csail.mit.edu/cchoi/vision4softhand/

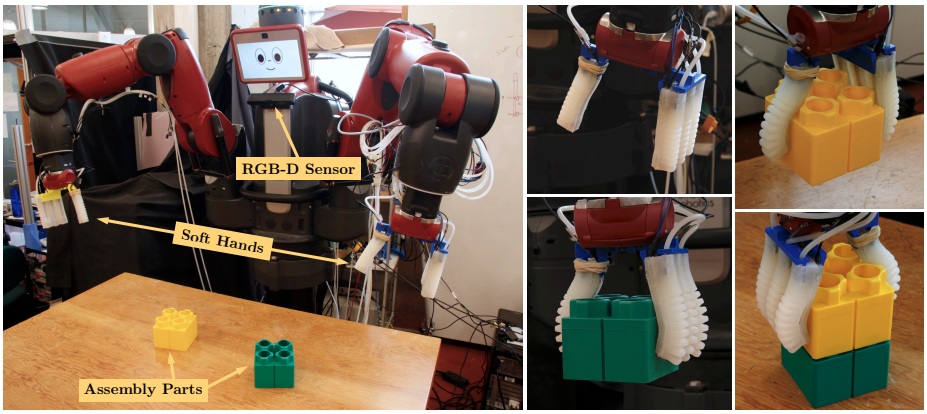
Abstract. In this paper, we present soft hands guided by an RGB-D object perception algorithm which is capable of localizing the pose of an object before and after grasping. The soft hands can perform manipulation operations such as grasping and connecting two parts. The flexible soft grippers grasp objects reliably under high uncertainty but the poses of the objects after grasping are subject to high uncertainty. Visual sensing ameliorates the increased uncertainty by means of in-hand object localization. The combination of soft hands and visual object perception enables our Baxter robot, augmented with soft hands, to perform object assembly tasks which require high precision. The effectiveness of our approach is validated by comparing it to the Baxter's original hard hands with and without the in-hand object localization.

Keywords: soft hands, soft gripper, in-hand object localization, pose estimation, robotic assembly, vision-guided manipulation

1 Motivation and Related Work

An important prerequisite for object manipulation is estimating the pose of an object and coping with the *uncertainty* of the pose estimates. Various sensing modalities, such as proprioception [1,2], visual exteroception [3,4], and contact/force sensing [5] have been employed. Visual sensing allows passive perception as it does not require contact, and is thus useful in the *pre-grasping* phase. Tactile, contact, force, and proprioceptive sensing modalities are useful when robots interact with objects in the *post-grasping* phase. The pose of a grasped object can be quite uncertain as the act of grasping tends to move the object and increase the uncertainty. Many prior works have combined vision and contact to decrease uncertainty [6,7,8,9,10,11,12,13].

Soft grippers are more compliant and easier to control than their hard counterparts [14,15,16,17]. The flexible materials of soft hands enable compliance with discrepancy between their belief space and the real environment; this compliance allows soft hands to be more tolerant of errors in the pose estimates of



(a) System setup (b) Soft hand

Fig. 1: **System overview.** Our system is composed of the Baxter robot augmented with two soft hands and an RGB-D sensor. Assembly parts are randomly placed on the table, so the positions and orientations of the parts are unknown. The RGB-D sensor localizes the parts on the table and inside the hand during the *pre-grasping* and *post-grasping* phases, respectively. The RGB channels are used for identification of the soft fingers, while the depth channel is employed for depth-based object localization.

objects. Softness, however, often reduces the confidence of the object state in the gripper since the pose of the object is more uncertain due to the flexibility of the soft fingers. In-hand object localization is thus needed for advanced object manipulations requiring accurate pose.

The goal of this paper is to develop a reliable object manipulation system using soft hands and visual pose feedback and to evaluate its effectiveness. Fig. 1a illustrates our system setup in which the Baxter robot is augmented with two soft hands and an RGB-D sensor. We use vision for localizing objects presented to the robot on a tabletop and then determining the pose of a grasped object in the hand. Fig. 1b shows one of our soft hands, which is composed of four pneumatically actuated fingers [2]. An RGB-D sensor is employed to localize objects in the workspace of the robot in the pre-grasping phase and to detect soft fingers and a grasped object in the post-grasping phase. Our approach does not rely on proprioceptive force sensing, yet it is capable of assembly operations requiring precision. To the best of our knowledge, this is the first attempt to use a vision-based object localization for soft hands capable of assembly tasks.

This paper is organized as follows. We explain the details of our technical approach in Section 2, wherein the problem statement and object localization algorithms are presented. Section 3 describes the experimental setup and results of two experimental tasks. Section 4 discusses concluding remarks and future work.

Algorithm 1: Pre-grasping Object Localization

Data: depth image $\mathcal{D}$, object models $\mathcal{M}$, in-plane rotation step Δ, minimum likelihood τ_l
Result: object poses $\widehat{\mathcal{X}}$, likelihoods $\widehat{\mathcal{L}}$, object indices $\widehat{\mathcal{O}}$

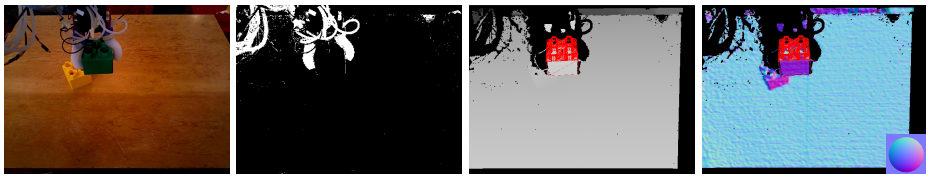
$\widehat{\mathcal{X}} \leftarrow \varnothing, \widehat{\mathcal{L}} \leftarrow \varnothing, \widehat{\mathcal{O}} \leftarrow \varnothing$ // initialize to empty sets
$\mathcal{S} \leftarrow$ `PlaneSeg(`$\mathcal{D}$`)` // segments objects $\mathcal{S}$ on a tabletop
for $s \in \mathcal{S}$ **do** // iterate over all segments $\mathcal{S}$
  $\mathcal{X} \leftarrow \varnothing, \mathcal{L} \leftarrow \varnothing, \mathcal{O} \leftarrow \varnothing$ // initialize to empty sets
  for $m \in \mathcal{M}$ **do** // consider all models $\mathcal{M}$
    $\mathcal{R} \leftarrow$ `In-planeRot(`m, Δ`)` // $\mathcal{R} \subset SO(3)$
    for $r \in \mathcal{R}$ **do**
      $\{\mathbf{X}, l\} \leftarrow$ `ICP(`$\left(\begin{smallmatrix} r & s \\ 0 & 1 \end{smallmatrix}\right), m, \mathcal{D}$`)` // init pose for ICP [18]
      if $l > \tau_l$ **then**
        $\mathcal{X} \leftarrow \mathcal{X} \cup \{\mathbf{X}\}$
        $\mathcal{L} \leftarrow \mathcal{L} \cup \{l\}$
        $\mathcal{O} \leftarrow \mathcal{O} \cup \{\mathcal{M}$`.index(`$m$`)`$\}$ // index() returns index of model m in $\mathcal{M}$
  $\{\hat{\mathbf{X}}, \hat{l}, \hat{o}\} \leftarrow \arg\max_{\mathcal{L}} \{\mathcal{L}, \mathcal{X}, \mathcal{O}\}$ // optimal estimate from the most $\mathcal{L}$
  $\widehat{\mathcal{X}} \leftarrow \widehat{\mathcal{X}} \cup \{\hat{\mathbf{X}}\}$
  $\widehat{\mathcal{L}} \leftarrow \widehat{\mathcal{L}} \cup \{\hat{l}\}$
  $\widehat{\mathcal{O}} \leftarrow \widehat{\mathcal{O}} \cup \{\hat{o}\}$

2 Technical Approach

Problem Statement: Given objects randomly placed a tabletop, we wish to enable a robot to grasp an object and connect it to another object on the table using vision as feedback. The robot is assumed to have soft hands. The objects have a known geometry $\mathcal{M}$. The stable grasp poses for the objects $\mathcal{X}_e^o \subset SE(3)$ and the extrinsic calibration of the RGB-D sensor $\mathbf{X}_c^w \in SE(3)$ are assumed to be known.

2.1 Pre-grasping Object Localization

The pre-grasping object localization estimates the poses of the objects on a planar table before the robot executes grasping, allowing for a pre-computed stable grasp to be realized. Algorithm 1 presents the pre-grasping object localization procedure. It takes the depth image $\mathcal{D}$ from an RGB-D sensor and a set of object models $\mathcal{M}$ as input, and then it returns a set of object poses $\widehat{\mathcal{X}} \subset SE(3)$ and their associated likelihoods $\widehat{\mathcal{L}} \subset \mathbb{R}^+$ and object indices $\widehat{\mathcal{O}} \subset \mathbb{N}$. We assume a table-top manipulation scenario where objects are placed on a planar table. This assumption allows the robot to segment foreground objects from the planar background. The function `PlaneSeg(`$\mathcal{D}$`)` first fits the plane model to the point cloud $\mathcal{D}$. Foreground point clouds of the objects are then clustered, and a set of



(a) RGB (b) Hand detection (c) Depth (d) Normal

Fig. 2: **Post-grasping object localization.** The hand regions (white in 2b) are estimated from a Gaussian naive Bayes classification on the hue and saturation channels from the RGB channels (2a) in the RGB-D sensor. The detected finger regions are then ignored in the depth-based object localization (2c, 2d). The red wireframes show the localized block part. (Best viewed in color)

Algorithm 2: Post-grasping Object Localization

Data: image $\mathcal{I}$, depth image $\mathcal{D}$, object model $m \in \mathcal{M}$, grasping pose $\mathbf{X}_e^o \subset \mathcal{X}_e^o$
Result: in-hand object pose $\widehat{\mathbf{X}}$, its likelihood $\widehat{l}$

1: $\mathcal{H} \leftarrow$ `DetectHandRegion`$(\mathcal{I}, \mathcal{D})$ `// detect hand region` $\mathcal{H}$
2: $\acute{\mathcal{D}} \leftarrow \mathcal{D} \cap \neg\mathcal{H}$ `// ignore hand region`
3: $\mathbf{X}_e^w \leftarrow$ `GetEndEffectorPose()` `// get EE pose via forward kinematics`
4: $\mathbf{X}_o^w \leftarrow \mathbf{X}_e^w (\mathbf{X}_e^o)^{-1}$ `// get object pose` $\mathbf{X}_o^w$ `in world coordinate frame`
5: $\{\widehat{\mathbf{X}}, \widehat{l}\} \leftarrow$ `ICP`$(\mathbf{X}_o^w, m, \acute{\mathcal{D}})$ `// ICP [18]`

center positions of the objects $\mathcal{S} \subset \mathbb{R}^3$ is estimated. The Iterative Closest Point (ICP) algorithm [18] is sequentially executed on each center position. As the unknown orientation of each object is constrained by the table, a set of in-plane rotations $\mathcal{R}$ with the step Δ is considered. Hence the initial pose for the ICP algorithm is set as $\left(\begin{smallmatrix} r & s \\ \mathbf{0} & 1 \end{smallmatrix}\right) \in SE(3)$ where $r \in SO(3)$ and $s \in \mathbb{R}^3$. Among the multiple ICP executions, the optimal pose estimate $\widehat{\mathbf{X}}$ with the most likelihood $\hat{l}$ is chosen for each point cloud cluster s. It is worth noting that the depth image $\mathcal{D}$ is in the camera coordinate frame, while the initial pose and the optimal pose estimate are with respect to the world coordinate frame. To transform between these two coordinate frames, the extrinsic calibration of the sensor $\mathbf{X}_c^w$ is estimated offline. The stable grasp poses for the objects $\mathcal{X}_e^o \subset SE(3)$ are also assumed to be known *a priori*, and thus a set of grasping poses for each object is accordingly calculated from the object pose estimates $\widehat{\mathcal{X}}$. Once this is done, the robot can be commanded to the desired pose and the grasp can be executed.

2.2 Post-grasping Object Localization

A challenge with visual in-hand object localization (IOL) is the occlusions caused by the grasping fingers. The performance of registration algorithms such as ICP is often deteriorated by occlusions. It is thus important to remove the regions of the fingers before running the registration algorithms. Traditionally, reasoning about finger locations has been done through model-based approaches where

an articulated shape model is rendered with the current state of joints [13]. However, the deformation of a soft finger is nonlinear so model-based approaches are difficult to derive and often too computationally intensive to use in real-time. Furthermore, the deformation of a soft finger varies depending on the grasped object shape and the contact points between the finger and the surface of the object.

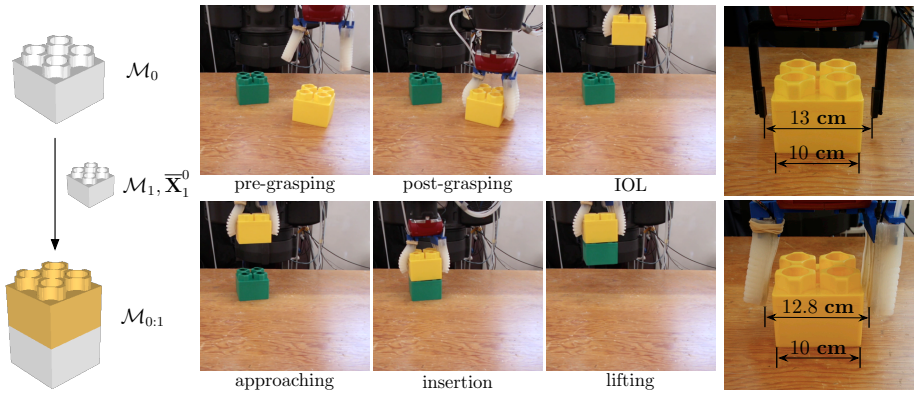
To address these issues, we adopted a data-driven approach in which a binary naive Bayes classifier [19] is trained to detect the fingers using the color data from the RGB-D sensor. Algorithm 2 presents the post-grasping object localization procedure. The RGB $\mathcal{I}$ and depth $\mathcal{D}$ images along with the grasped object model $m \in \mathcal{M}$ and grasped pose $\mathbf{X}_e^o \in \mathcal{X}_e^o \subset SE(3)$ are given, and the algorithm returns the refined object pose $\widehat{\mathbf{X}}$ with its likelihood $\widehat{l}$. The function `DetectHandRegion(`$\mathcal{I}$`,` $\mathcal{D}$`)` detects the soft hand regions $\mathcal{H}$ via the naive Bayes classifier. To train the classifier, $\mathcal{D}$ from the sensor is employed to segment the soft fingers and the background, and the color distributions of the soft finger and background regions are used as positive and negative training data, respectively. We adopted the HSV color space and used H (hue) and S (saturation) channels for better invariance to brightness changes. The color distributions for both the positive and the negative examples are modeled by the mixture of Gaussians. The white area in Fig. 2b shows the soft hand regions $\mathcal{H}$ detected from the trained naive Bayes classifier. The region $\mathcal{H}$ is then used as an erasing mask for $\mathcal{D}$ so that the in-hand object localization is done on the depth image without the hand regions $\acute{\mathcal{D}}$. The initial object pose $\mathbf{X}_o^w$ in the world coordinate frame w is estimated from the end effector pose $\mathbf{X}_e^w$ and the grasping pose $\mathbf{X}_e^o$. Fig. 2c and 2d show the red wireframes of the object model m with the refined pose $\widehat{\mathbf{X}}$ on the depth and surface normal images, respectively.

3 Experiments

We augmented the Baxter with two hands of four soft fingers each as shown in Fig. 1a. We tasked the robot to pick up one block with one hand then connect it to the other block on the tabletop. Fig. 3 shows the block assembly procedure and the step-wise stages of the assembly. Once the Baxter grasps the block object, it re-localizes the block in the hand with the in-hand object localization (IOL). It then approaches to the top of the second block on the table and connects the grasped block to the block on the table. To make sure that the blocks are well inserted together, it lifts the assembled blocks. If the two blocks are lifted together, the assembly task is *successful*, otherwise *unsuccessful*. The rightmost column of Fig. 4 shows some successful and unsuccessful examples.

To investigate the effectiveness of the soft hands and the post-grasping object localization, we compare the two aspects: hard gripper vs. soft gripper, and with and without the IOL. There are thus four different configurations considered in this experiments:

1. The hard gripper without the IOL (**H**): This configuration is using the original hard gripper of the Baxter, and not using the post-grasping object local-



(a) Procedure (b) Block assembly with the IOL (c) Hand size

Fig. 3: **Block assembly.** The assembly procedure (3a) is to connect the block $\mathcal{M}_1$ to the other block $\mathcal{M}_0$ with the relative pose $\overline{\mathbf{X}}_1^0 \in SE(3)$. The sequence of figures (3b) shows that the Baxter grasps the block with its left soft hand and inserts it on the other block. The distance between fingers for both hand types is about 13**cm** (3c).

ization. This configuration serves as a baseline for comparative experiments in the following subsections.
2. The hard gripper with the IOL (**HI**): It uses the original hard gripper, but it localizes the object after grasping.
3. The soft gripper without the IOL (**S**): It uses the soft hand instead of the hard one, but does not localize objects after grasping.
4. The soft gripper with the IOL (**SI**): This configuration uses the soft hand and the IOL to localize the object in the hand. It is the configuration of our system.

We investigate the effectiveness of soft hands by comparing the hard hands (**H**, **HI**) and the soft hands (**S**, **SI**) respectively. The effectiveness of the IOL can also be seen by comparing the performance with (**HI**, **SI**) and without (**H**, **S**) the IOL. We compare these four configurations in the two evaluation scenarios as follows:

1. Evaluation with respect to artificial Gaussian noise in object pose
2. Evaluation of the complete system.

Detailed experimental settings are explained in the subsequent sections.

3.1 Robustness to Object Pose Noise

The purpose of this experiment is to compare the robustness and accuracy of the object manipulation with respect to the noise in object pose. The considered manipulation tasks include *grasping* and *insertion*. The success of such manipulation depends on object pose estimates which are calculated by the pre- and post-grasping object localization. The successful manipulation also depend on

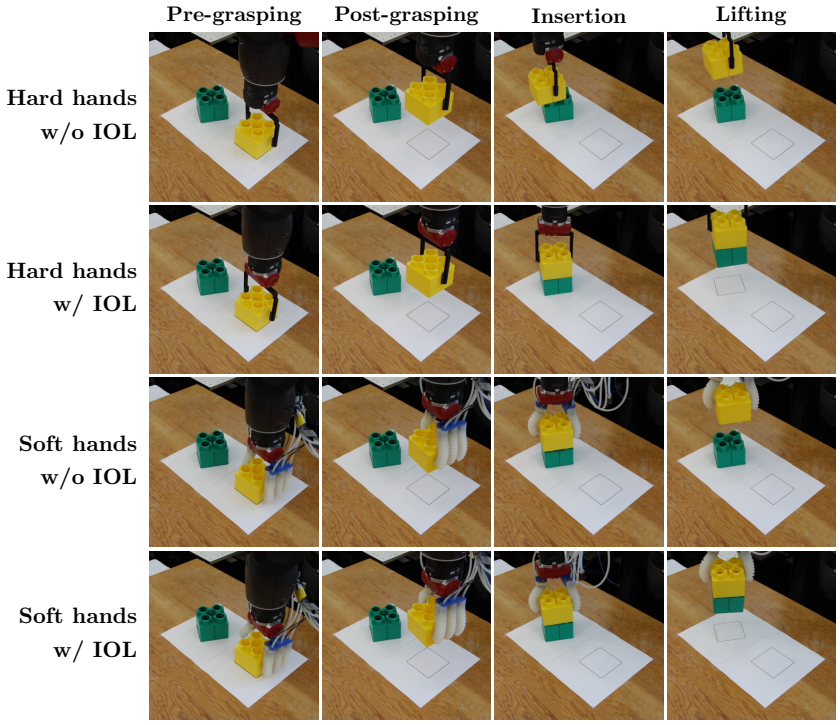


Fig. 4: **A grasping and insertion example of the four configurations.** The same Gaussian noise was added to the four configurations, yet the outcome of the insertion task is different. The added noise in this experiment was not enough to cause grasping failure, but it does cause insertion failure unless the IOL is used. Thus, using the IOL makes insertion more robust to pose errors.

the arm trajectories solved from the motion planning algorithm. In order to evaluate the differences between the four configurations (**H**, **HI**, **S**, and **SI**), we minimize extraneous sources of error by maintaining a consistent configuration. As shown in Fig. 4, we fix the poses of the two blocks on the table and the same pose estimates are used for the four configurations. In order to ensure consistent object locations, two sheets of white paper were affixed to the table and the blocks were carefully aligned before each trial. The robot is tasked to pick up the left block and connects it to the right block. In order to get the beginning poses of the blocks, we execute the pre-grasping object localization multiple times (100 times in our experiment) and calculate the mean of the multiple pose estimates.[1] It turns out that the accuracy of our pre-grasping object localization algorithm shows sub-millimeter and sub-degree uncertainties in translation and

[1] The pre-grasping object localization was run 100 times. For each pose estimate $\mathbf{X}_i = \begin{pmatrix} \mathbf{R}_i & \mathbf{t}_i \\ \mathbf{0} & 1 \end{pmatrix} \in SE(3)$, the standard deviation of the translation is calculated from the arithmetic mean of the translation $\bar{\mathbf{t}} = \sum_i \mathbf{t}_i / N$ where N is the number of samples (i.e. $N = 100$). While the translation vectors $\mathbf{t}_i$ are in Euclidean space, the rotation matrices are in the special Orthogonal group $SO(3)$. We thus need to take

rotation, respectively.[2] As the uncertainty of the pre-grasping localization is not significant, we add artificial noise to the object poses for the following evaluation.

To evaluate the effectiveness of the four configurations with respect to the uncertainty in object pose, we add artificial Gaussian noise to the mean pose estimates. For fair comparison, we generate a series of Gaussian noise in the object pose and use the same noise series with each of the four configurations. As the block objects are on the table, we add Gaussian noise on the plane by adding in $\mathbf{x}$, $\mathbf{y}$, and $\boldsymbol{\theta}$:

$$\mathbf{X}_i = \overline{\mathbf{X}}\widetilde{\mathbf{X}}_i \tag{1}$$

where $\mathbf{X}_i \in SE(3)$ is the noise-perturbed pose and $\overline{\mathbf{X}} \in SE(3)$ is the noise-free pose obtained by the mean of the multiple pose estimates from the pre-grasping object localization. The noise $\widetilde{\mathbf{X}}_i \in SE(3)$ is sampled from Gaussian distributions as follows:

$$\widetilde{\mathbf{X}}_i = \begin{pmatrix} \mathbf{R}_z(\theta) & \mathbf{t} \\ \mathbf{0} & 1 \end{pmatrix} \tag{2}$$

where $\mathbf{R}_z(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \in SO(3), \theta \sim \mathcal{N}(0, \sigma_r^2)$ is the 3D rotation along the $\mathbf{z}$-axis and $t_x, t_y \sim \mathcal{N}(0, \sigma_t^2)$ where $\mathbf{t} = (t_x, t_y, 0)^\intercal$ is the translation from the center of the noise-free object pose.[3] In our experiment, we used $\sigma_t = 20\mathbf{mm}$ and $\sigma_r = 20°$. Fig. 4 shows a grasping and insertion trial of the four configurations with the same Gaussian noise. The Gaussian noise for this trial is -13.5**mm** in $\mathbf{x}$-axis, 0.6**mm** in $\mathbf{y}$-axis, and $-12.7°$ along $\mathbf{z}$-axis. The error is small enough that all configurations are successful in grasping the left yellow block. For the insertion task, however, we can notice that the configurations with the IOL (**HI**, **SI**) are successful, while those without the IOL (**H**, **S**) are not accurate enough for the task. The insertion task requires much tighter tolerance than the grasping task, and hence enhancing the uncertainty of the object pose in the hand is crucial for such insertion task.

Fig. 5 presents the plots of the grasping and assembly results with respect to Gaussian noise in the object pose. By comparing left and right columns, we notice a significant improvement in the success rate of the assembly operation in both hard and soft hands. As the IOL refines the pose of the objects, the uncertainty of the object in the hand is significantly reduced, and hence the

special consideration into this $SO(3)$ space. It is well known that the geodesic metric on $SO(3)$ is the angle between two rotation matrices $d(\mathbf{R}_1, \mathbf{R}_2)$ and the valid mean of a set of rotation matrices can be estimated from the geometric mean [20]. The standard deviation in the angle is calculated from the geometric mean.

[2] The standard deviations of the (x, y, z) translation in the 100 object pose estimates are $(0.14, 0.46, 0.28)$ **mm** and $(0.18, 0.38, 0.20)$ **mm** for the left and right blocks respectively. The standard deviations of the angle distance between each rotation matrix in $SO(3)$ and the mean of the rotation matrices are $0.26°$ and $0.17°$ for the left and right blocks respectively.

[3] The rotational axis of the block object model is the $\mathbf{z}$-axis.

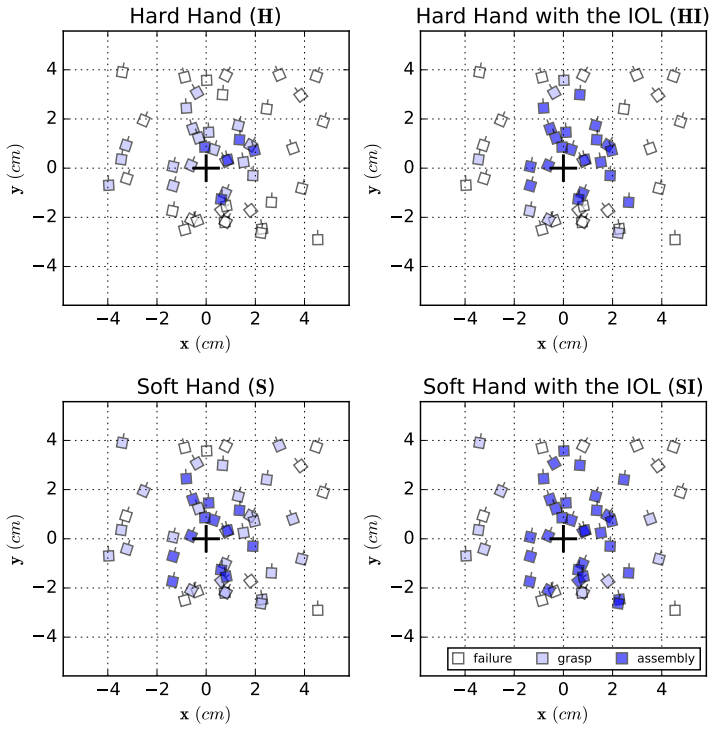


Fig. 5: **Plots of the results with respect to Gaussian noise in the object pose.** Each of the four configurations was executed 50 times with a series of pre-generated Gaussian noise in the block object pose. Each square represents one trial in which its location and orientation depict the Gaussian noise in the translation ($\mathbf{x}$ and $\mathbf{y}$) and the orientation (θ). A white square means unsuccessful to grasp the object; a lightly shaded blue square represents successful grasping but failure of assembly, while the dark blue square shows successful grasping and assembly. The symbol + represents the origin of the object coordinate frame.

number of successful trials is noticeably increased. Another direction to compare is the hard hand (first row) and soft hand (second row). If we look at the left column showing the results without the IOL, we see a noticeable improvement in both grasping and assembly tasks. It demonstrates the adaptability of the soft hands with respect to pose noise. In the lower plots, even if the block is off about 4**cm** in both axes, the soft gripper can grasp the noise-perturbed object, while the hard gripper shows less promising results. The compliance of the soft gripper plays an important role even with the IOL, as can be seen in the right column. These experimental results clearly show the effectiveness of the compliant soft grippers, which are adaptable to the noisy pose estimates, as well as of the IOL.

Table 1 presents the results of the four configurations in terms of the numbers of each result and the success rates. If we compare the hard hands (**H**, **HI**) with the soft hands (**S**, **SI**), we notice a significant improvement in the success rate of

Table 1: Success rates for 50 trials of the Gaussian noise experiment.

Measure	Hard Hand		Soft Hand	
	¬IOL (**H**)	IOL (**HI**)	¬IOL (**S**)	IOL (**SI**)
# of Failure	27	23	11	11
# of Grasping	18	7	26	9
# of Assembly	5	20	13	30
Successful Grasping†	46%	54%	78%	78%
Successful Assembly†	10%	40%	26%	60%

† The success rate of grasping considers both '# of grasping' and '# of assembly'.

grasping. The success rate of grasping for hard gripper is about 50% on average, while that for soft gripper is almost 80%. Even when the pose estimate of each object was perturbed, the soft gripper tends to adapt to the pose error due to the flexibility in the soft materials. Another notable difference is the success rate of the block assembly between the configurations with and without the IOL. For the hard gripper, the IOL enables the Baxter to assemble the blocks correctly and thus the success rate is four times higher than without the IOL. Similar effects can be found in the soft gripper, where **SI** shows 60% success rate while **S** is successful about one fourth. Running the IOL reduces uncertainty in the pose of the in-hand object, and thus it increases the success rate of the block assembly task which requires a tight tolerance.

3.2 Evaluation of the Complete System

In the second evaluation, we compare the four configurations in less constrained settings. The setup is similar to the experiment in Section 3.1, but the two blocks are randomly placed on the table and the robot randomly picks one of the blocks and connects it to the other block. The Gaussian noise is not added to the pose estimate from the pre-grasping object localization. This setting is to evaluate the complete system with uncertainties from the pose estimate of the object localization algorithms, planning trajectories, robot calibration, etc.

Table 2 shows the success rates of the block assembly on the table in the four configurations. As we explained in Section 3.1, our pre-grasping object localization returns sub-millimeter accuracy in translation and sub-degree accuracy in rotation. Without the additional Gaussian noise, we notice that grasping the block object is not a challenging problem for both hard and soft hands. It is, however, still challenging for the assembly task. When the hard hand is considered without the IOL, the success rate is only 41%. But the same hand with the IOL improves the success rate of the assembly task to 66%, which is more than 20% improvement. A similar trend can be observed in the soft hand configuration. Without the IOL it is successful in 72% of trials, but using the IOL enables it to succeed in over 90% of trials. If we compare hard and soft hands, we notice that there is about 30% improvement when using soft hands (41% to 72% and 66% to

Table 2: Success rates for 100 trials of the complete system experiment.

Measure	Hard Hand		Soft Hand	
	¬IOL (**H**)	IOL (**HI**)	¬IOL (**S**)	IOL (**SI**)
Successful Grasping	100%	100%	100%	100%
Successful Assembly	41%	66%	72%	92%

92%). These results therefore confirm both the effectiveness of using adaptable, flexible soft hands and of using the IOL. Together, they can yield successful manipulation in challenging scenarios.

4 Conclusion

We proposed an object manipulation approach which provides flexibility through compliant soft hands and dependable accuracy using vision-based localization algorithms. The color and depth channels were effectively employed for soft finger segmentation and object localization, respectively. The object pose in the soft hands is prone to be uncertain due to the flexible deformation of the soft hands. Nevertheless, our in-hand localization approach is effective in mitigating this problem. The compliance of the soft hands is adaptable to the uncertainty in object pose, and thus it is effective for manipulation tasks which require a tight tolerance.

For future work, we would like to extend this approach to dual-arm manipulation which is capable of more sophisticated manipulation such as assembling two object parts with two hands in air. This dual-hand manipulation doubles the uncertainties in both hands and objects. We anticipate that this manipulation will be a challenging scenario for which our in-hand object localization and compliant fingers can be very advantageous.

Acknowledgement

This work was supported by The Boeing Company. The support is gratefully acknowledged.

References

1. Mason, M.T., Rodriguez, A., Srinivasa, S.S., Vazquez, A.S.: Autonomous manipulation with a general-purpose simple hand. International Journal of Robotics Research **31**(5) (2012) 688–703
2. Homberg, B.S., Katzschmann, R.K., Dogar, M.R., Rus, D.: Haptic identification of objects using a modular soft robotic gripper. In: Proc. IEEE/RSJ Int'l Conf. Intelligent Robots Systems (IROS). (September 2015) 1698–1705

3. Collet, A., Martinez, M., Srinivasa, S.S.: The MOPED framework: Object recognition and pose estimation for manipulation. International Journal of Robotics Research **30**(10) (2011) 1284–1306
4. Choi, C., Christensen, H.I.: Robust 3D visual tracking using particle filtering on the Special Euclidean group: A combined approach of keypoint and edge features. International Journal of Robotics Research **31**(4) (April 2012) 498–519
5. Bimbo, J., Luo, S., Althoefer, K., Liu, H.: In-hand Object Pose Estimation using Covariance-Based Tactile To Geometry Matching. IEEE Robotics and Automation Letters **1**(1) (2016)
6. Allen, P.K.: Integrating vision and touch for object recognition tasks. International Journal of Robotics Research **7**(6) (1988) 15–33
7. Hebert, P., Hudson, N., Ma, J., Burdick, J.: Fusion of stereo vision, force-torque, and joint sensors for estimation of in-hand object location. In: Proceedings of IEEE International Conference on Robotics and Automation (ICRA). (May 2011) 5935–5941
8. Zhang, L.E., Trinkle, J.C.: The application of particle filtering to grasping acquisition with visual occlusion and tactile sensing. In: Proceedings of IEEE International Conference on Robotics and Automation (ICRA), IEEE (2012) 3805–3812
9. Ilonen, J., Bohg, J., Kyrki, V.: Fusing visual and tactile sensing for 3-D object reconstruction while grasping. In: Proceedings of IEEE International Conference on Robotics and Automation (ICRA). (2013) 3547–3554
10. Bimbo, J., Seneviratne, L., Althoefer, K., Liu, H.: Combining touch and vision for the estimation of an object's pose during manipulation. In: Proc. IEEE/RSJ Int'l Conf. Intelligent Robots Systems (IROS). (November 2013) 4021–4026
11. Chalon, M., Reinecke, J., Pfanne, M.: Online in-hand object localization. In: Proc. IEEE/RSJ Int'l Conf. Intelligent Robots Systems (IROS). (November 2013) 2977–2984
12. Guler, P., Bekiroglu, Y., Gratal, X., Pauwels, K., Kragic, D.: What's in the container? Classifying object contents from vision and touch. In: Proc. IEEE/RSJ Int'l Conf. Intelligent Robots Systems (IROS), IEEE (2014) 3961–3968
13. Schmidt, T., Hertkorn, K., Newcombe, R., Marton, Z., Suppa, M., Fox, D.: Depth-Based Tracking with Physical Constraints for Robot Manipulation. In: Proceedings of IEEE International Conference on Robotics and Automation (ICRA). (2015)
14. Deimel, R., Brock, O.: A compliant hand based on a novel pneumatic actuator. In: Proceedings of IEEE International Conference on Robotics and Automation (ICRA), IEEE (2013) 2047–2053
15. Stokes, A.A., Shepherd, R.F., Morin, S.A., Ilievski, F., Whitesides, G.M.: A Hybrid Combining Hard and Soft Robots. Soft Robotics **1**(1) (July 2013) 70–74
16. Deimel, R., Brock, O.: A novel type of compliant and underactuated robotic hand for dexterous grasping. International Journal of Robotics Research **35**(1-3) (January 2016) 161–185
17. Galloway, K.C., Becker, K.P., Phillips, B., Kirby, J., Licht, S., Tchernov, D., Wood, R.J., Gruber, D.F.: Soft Robotic Grippers for Biological Sampling on Deep Reefs. Soft Robotics (2016)
18. Besl, P.J., McKay, N.D.: A method for registration of 3-D shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence (1992) 239–256
19. Murphy, K.P.: Machine Learning: A Probabilistic Perspective. The MIT Press (August 2012)
20. Moakher, M.: Means and averaging in the group of rotations. SIAM Journal on Matrix Analysis and Applications **24**(1) (2003) 1–16