# Hanging a T-Shirt: A Step towards Deformable Peg-in-Hole Manipulation with Multimodal Tactile Feedback

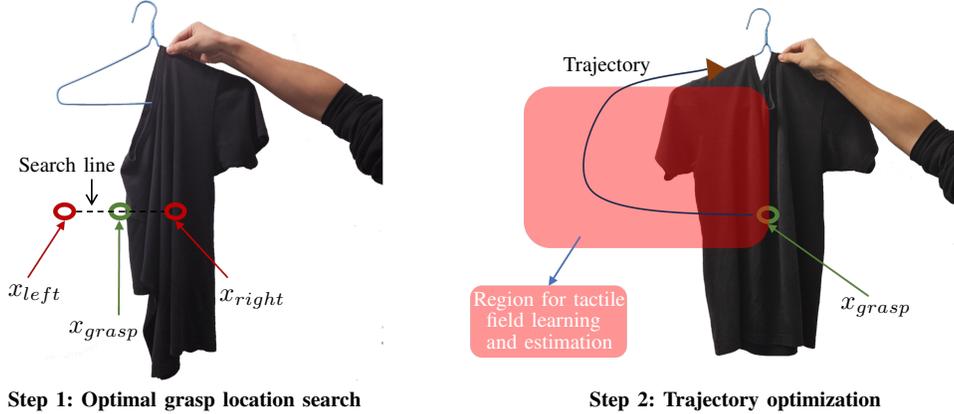Yipai Du[1], Shoaib Aslam[1,2], Michael Yu Wang[3], *Fellow, IEEE*, Bertram E. Shi[1], *Fellow, IEEE*

Fig. 1: The two steps of the task. Left: the first step is to search for the grasp point so it grasps at the edge of the T-shirt. Right: the second step is to plan and execute a force-aware trajectory to enclose the neckline.

*Abstract*— **Peg-in-hole manipulation has been a long-standing problem in robotics due to its broad application in both domestic and industrial domains. Due to advances in perception and modeling, deformable object manipulation is attracting increasing attention. This paper focuses on the intersection of these problems, where a hole must be deformed to allow peg entry. A common domestic application of this task is to pass a clothes hanger through a T-shirt's neckline to hang it. We demonstrate that the complexity of the problem can be reduced by using multimodal tactile feedback from a GelSight sensor. High-resolution contact detection helps to localize grasp to suitable locations on the T-shirt. Using tactile feedback to track a force trajectory, our algorithm can manipulate T-shirts of varying size and stiffness so that their neckline encloses the hanger. Our experimental results show consistency with the theoretical analysis. We anticipate that our proposed method will be more broadly applicable to other problems requiring simultaneous estimation and execution of force trajectory for elastic objects.**

## I. INTRODUCTION

Deformable object manipulation is required in many scenarios, such as industrial wire harnesses, fruit picking, medical surgery, and home care. Despite its importance, it has been less investigated historically than rigid object manipulation, mainly due to the complicated modeling, perception, and control it requires [1].

Peg-in-hole manipulation has also been studied extensively due to its promising applications in industrial autonomous assembly[2] and for domestic tasks [3].

This study examines the intersection of deformable manipulation and peg-in-hole problems: cloth hanging. The manipuland is deformable. The motion is peg-in-hole. It is a common everyday task where robots may be able to assist individuals with motor impairments. Here, we assume that the hanger is already half inserted in the T-shirt and held by one hand. While this is not easy, prior works have addressed this in the context of robotic dressing assistance [3], [4]. Our focus is to grasp the other edge of the collar and manipulate it over the uncovered end of the hangar (Fig. 1). Humans commonly use tactile feedback in this task to estimate the state of the T-shirt. This task is demanding if only visual feedback is available. The deformable object might be occluded. Even without occlusion, it is computationally demanding to estimate the state of the stretched T-shirt using state-of-the-art vision techniques. Visual feedback is not even necessary, as humans can perform this task in the dark.

Here, we take advantage of the multimodal (RGB/marker, i.e., texture/shear force) feedback from the GelSight tactile sensor [5] to alleviate the complexities in perception and estimation. Here, we use the sensor's capability to measure the contact texture and the shear load. Visual cues are only used in the initial stage to provide the range for grasp search. For all remaining steps: grasping and manipulation of the T-shirt's neckline, the tactile signal guides the action in real-time.

The main contributions of this work are summarized as follows:

- Contact-detection-based grasp location determination to reduce the DOF of the cloth
- Shear force field estimation and force trajectory generation

[1] The Hong Kong University of Science and Technology, Hong Kong.
[2] Department of Mechanical and Mechatronis Engineering, University of Engineering and Technology Lahore (Faisalabad Campus), Pakistan.
[3] School of Engineering, Great Bay University, Songshan Lake, Dongguan, Guangdong, China.
Correspondence to `yipai.du@connect.ust.hk`

- Robust t-shirt enclosure and hanging on the hanger.

After summarizing related studies in Sec. II, our approach is presented in detail in Sec. III. Experiments and validation of our method are given in Sec. IV followed by the conclusion and future works in Sec. V.

## II. RELATED WORKS

Manipulating 2D deformable objects through tactile sensor feedback during insertion is complex. Despite attracting attention from the robotics community, deformable object manipulation remains a challenge for robots [6]–[8]. Manipulating 1D deformable objects, a single linear component, involves handling cable and rope-like items [9]–[11]. Handling items like clothing and fabrics establish the class of 2D deformable objects, as examined in recent studies on fabric smoothing [12], [13], involving quality evaluation based on coverage. Further, garment unfolding [14] and assistive dressing [3] involve tasks utilizing 2D deformables. Lastly, handling plastic bags [6] extends the task to 3D deformable objects.

Peg-in-hole manipulation is a vibrant research field with applications in robot autonomous assembly [2], but less attention has been paid to the deformable objects in this area [8].

This study focuses on how tactile sensing can significantly enhance the robot's capabilities in handling deformable objects during peg-in-hole manipulation tasks. Recent advancements in high-resolution vision-based tactile sensing [5] have enabled the seamless integration of tactile feedback into complex manipulation scenarios that demand contact-rich local grasp feedback. Many existing methods have utilized visual and tactile sensing to address cloth material perception [15], grasping and edge-sliding [16], garment unfolding and dressing [14], [17], and plastic bag opening [8].

Simulating the closed loop deformable objects with the tactile imprints is challenging [18]. Most prior studies considered free deformable, rather than stress deformable, regions, except for Alberta *et al.*, who attempted to encode the elasticity of textiles [19]. However, their work did not use it for closed-loop control. Inspired by Hanai *et al.* who predicted the contact force distribution using visual feedback for rigid object manipulations [20], we believe predicting and tracking the elastic force will be helpful for deformable manipulations. In this work, we aim to address this gap by leveraging multimodal tactile feedback from the GelSight sensor for motion planning. Its measurements are directly from the contact surface, which enables accurate grasp point identification and shear force tracking to successfully hang a deformable closed loop collar onto a rigid hanger.

## III. APPROACH

The task starts with the configuration in which half of the hanger is inserted into the T-shirt so the T-shirt with the inserted part of the hanger is already fixed (with potentially another robot hand) as in Fig. 1. In some extreme cases, the T-shirt may be crumpled, so a flinging and shaking action is applied at the initial stage inspired by [8], [21]. The objective

for the robot is to find a good grasp location, grasp the T-shirt, and raise it so the other half of the hanger can get into the T-shirt through the collar. We name this step the neckline enclosing action. It represents a deformable peg-in-hole problem, in which an elastic hole is grasped and intentionally stretched for a proper deformation to get the peg inside.

This section first shows that trajectories without feedback cannot achieve this action robustly from a geometry and elastic deformation perspective. Then, we propose to mitigate this issue by exploiting the tactile feedback. This problem is formulated as a tactile estimation and trajectory optimization problem, and by alternating between these two problems on the fly, the desired force is produced on the trajectory. Upon following the desired force profile, the task can be accomplished with the appropriate deformation of the T-shirt.

### A. *Fundamental Issues in Trajectories without Force Feedback*

While position-based trajectories can succeed in enclosing the neckline in some cases, it is hard to achieve robustness in different scenarios. Although we can determine proper grasp location horizontally with trial grasps (Sec. III-C), the vertical grasp location may vary due to changes in the T-shirt or its configuration, etc. Fig. 2 illustrates one example when the actual grasp location is higher than the desired position, resulting in a vertical trajectory shift. Point $A$ is the desired grasp point, and point $B$ is the trajectory endpoint. Point $A'$ is the actual grasp point, and the trajectory is consequently shifted and ends at point $B'$. Let us define the strain (deformation) for the desired case as $dx$ and the actual case $dx'$, then
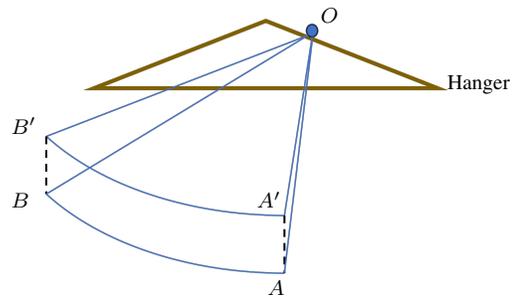
$$dx = OB - OA$$
$$dx' = OB' - OA'. \tag{1}$$



Fig. 2: A geometric illustration of the vertical uncertainty of grasp locations. Trajectory AB: the nominal case. Trajectory A'B': the actual case. O: the simplified fixed contact point on the hanger.

For the 2 triangles $\triangle$OAA' and $\triangle$OBB', it can be proved that $OA - OA' > OB - OB'$ (see Appendix). Therefore,

$$dx = OB - OA$$
$$< OB' - OA' \tag{2}$$
$$= dx'.$$

Moreover, when the grasp location is higher, in addition to larger strain, the natural length of the deformable region is shorter, which means the T-shirt will react with a higher stiffness. Hence, we can conclude that the elastic force on trajectory A'B' is higher than on the original trajectory AB. When the force becomes too big, it is unnecessary and can potentially damage the T-shirt and the hanger. When the grasp location is lower, the force is insufficient, making the task unsuccessful (see Fig. 3).
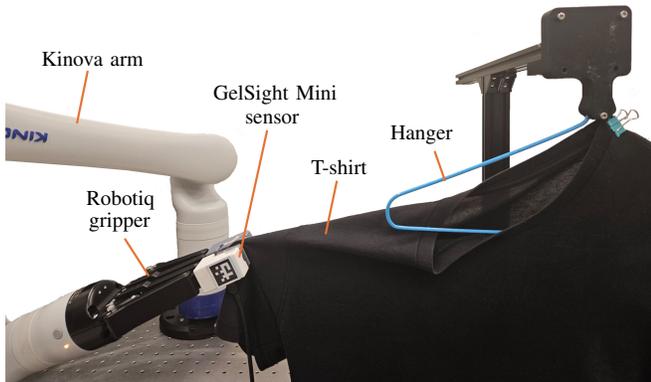


Fig. 3: An example when the grasping location is 2cm lower than desired. The dragging force is insufficient, so the T-shirt collar is not wide enough to fit the hanger.

Therefore, feedback signals are needed to guide the trajectory adjustment to accomplish the task. Instead of relying on visual feedback, which usually requires adapting to different types of T-shirts with varying cloth and colors in different shape configurations, we hope to rely purely on tactile feedback, which has much less variation and lower dimensionality. This characteristic of tactile signals can contribute to better generalization for the control policies across different scenarios.

### B. Problem Formulation

The problem of neckline enclosing can be divided into two stages. In the first stage, the robot approaches the T-shirt and looks for a proper gripper position $\mathbf{x}_{grasp}$ to grasp the T-shirt for subsequent actions. In the second stage, after grasping the T-shirt, the robot plans a cartesian trajectory $\{\mathbf{x}_i\}_{i=1}^T$ such that when the trajectory reaches the end and the gripper releases, the T-shirt will be readily hanging on the hanger. We assume the hanger position and orientation are known because it is grasped by the robot's other hand. Even if this is not the case, getting hanger orientation precisely from an RGBD camera is possible. Therefore, the plane defined by the flat hanger is assumed to be known and fixed. Hence, the robotic gripper needs only to move on this plane. This means the planning outcomes for both stages $\mathbf{x}_{init}$ and $\{\mathbf{x}_i\}_{i=1}^T$ are in $\mathbb{R}^2$.

To plan a successful trajectory $\{\mathbf{x}_i\}_{i=1}^T$, we require in addition a demonstration trajectory $\{\mathbf{x}_i^d\}_{i=1}^T$ from human expert and its corresponding sensor measured shear force $\{\mathbf{F}_i^d\}_{i=1}^T$ available. This demonstration trajectory is informative in the sense that it contains the overall shape of the desired trajectory and the proper force on it, hence we exploit this one demonstration to finish the task. We use the mean marker displacement over the 7x9 marker grid from the GelSight sensor as a proxy for the shear force. The sensor's shear deformation directly correlates with the shear force. A mapping from the marker displacement to shear force can be established to obtain the true force [5], but it is not required here. The aim of planning the trajectory $\{\mathbf{x}_i\}_{i=1}^T$ is to find a trajectory that produces a force profile that mimics the demonstration, i.e., closest to $\{\mathbf{F}_i^d\}_{i=1}^T$. This intuition comes from the experience when humans execute the enclosing motion: the force exerted by the T-shirt on the hand indicates how to move the hand to accomplish the task. In this case, positional trajectories are more diversified and can be less generic than force trajectories. Hence, reproducing the force profile is essential for task completion and makes it possible to generalize across varying environments with potentially different T-shirts. In addition, our method also requires a dataset of position-force pairs $\mathbf{D} = \{(\mathbf{x}_m^D, \mathbf{F}_m^D)\}_{m=1}^M$ as the force context. From the dataset, a differentiable model is learned offline and refined online using past history feedback. This dataset is collected with the same grasp location of only a single T-shirt. Still, the encoded information can help the robot understand the task setup and make force predictions in novel scenarios, i.e., different T-shirts and grasp locations.

The graphical illustration of the two stages of enclosure is shown in Fig. 1. We believe this problem formulation of planning a trajectory to mimic a force profile is beneficial to neckline enclosure and potentially helpful for stress deformable manipulation problems in general.

### C. Grasp Location Identification

From human experience, a good grasp location for neckline enclosing is vertically around the middle of the T-shirt and horizontally close to the edge of the T-shirt. Grasping at an inner region rather than the edge may fail to open the collar sufficiently. Inspired by this observation, we fixed the vertical grasp height and reduced the grasp location problem to a 1D horizontal search problem. Algorithm 1 details how to search for the edge of the T-shirt as an initial grasp location. Specifically, it relies on the estimate of the approximate range $\mathbf{x}_{left}$ and $\mathbf{x}_{right}$ and iteratively shrinks the range based on similarity-based grasp detection to check if there is cloth grasped between the fingers. The values of $\mathbf{x}_{left}$ and $\mathbf{x}_{right}$ can be easily extracted from visual information (e.g., semantic segmentation) or based on the known hanger geometry settings. The advantage of this detection method is that it requires no anticipation over what will be grasped but is entirely based on the information when the empty fingers close. The grasp detection of unknown objects like thin cloth is complex for traditional low-resolution tactile sensors but is feasible with the high-resolution tactile images from the GelSight sensor. It restricts the grasp location to the T-shirt edge and reduces the uncertainties of the downstream trajectory following task.

**Algorithm 1:** Grasp Location Identification

**Input:** Location guaranteed to grasp the T-shirt
$\mathbf{x}_{right}$,
location guaranteed to have no T-shirt grasped $\mathbf{x}_{left}$,
gripper close width $w$,
gripper width variation $\delta$.

**Output:** Grasp location $\mathbf{x}_{grasp}$.

1 Close the gripper to $w - \delta$, record the tactile image $T_{w-\delta}$;
2 Close the gripper to $w$, record the tactile image $T_w$;
3 Close the gripper to $w + \delta$, record the tactile image $T_{w+\delta}$;
4 $threshold = \min\left(ssim(T_w, T_{w-\delta}), ssim(T_w, T_{w+\delta})\right);$
5 Release the gripper;
6 **while** $distance(\mathbf{x}_{left}, \mathbf{x}_{right}) > 1cm$ **do**
7    Gripper goes to $\frac{\mathbf{x}_{left} + \mathbf{x}_{right}}{2}$;
8    Set gripper width to $w$, record the tactile image $I_t$;
9    **if** $ssim(I_t, T_w) > thsreshold$ **then**
10       $\mathbf{x}_{left} = \frac{\mathbf{x}_{left} + \mathbf{x}_{right}}{2}$;
11    **else**
12       $\mathbf{x}_{right} = \frac{\mathbf{x}_{left} + \mathbf{x}_{right}}{2}$;
13    **end**
14    Release the gripper;
15 **end**
16 $\mathbf{x}_{grasp} = \mathbf{x}_{right}$;
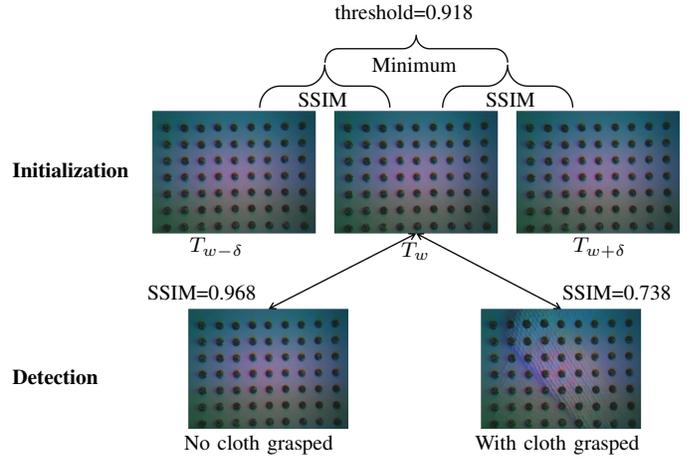17 **return** $\mathbf{x}_{grasp}$



Fig. 4: The tactile images captured by the GelSight sensor during threshold initialization and contact detection.

$\mathbf{A}_t \in \mathbb{R}^{2\times 2}$ is the affine transformation matrix and $\mathbf{b}_t \in \mathbb{R}^{2\times 1}$ is the offset matrix. Eq. 3 is a linear least squares problem; hence, $\mathbf{A}$ and $\mathbf{b}$ can be solved analytically. The optimization seeks a linear transformation of the nominal tactile field that best matches the measured force history up to the current time step. Then, in the next step, we can optimize the trajectory in this estimated tactile field to produce the desired force profile.

### E. Tactile Trajectory Generation

With the estimated tactile field $N_t$, the target trajectory can be generated with the following optimization problem.
**The trajectory generation problem**:

$$\mathbf{M}_{t+1}, \mathbf{w}_{t+1} = \underset{\mathbf{M}, \mathbf{w}}{\operatorname{argmin}} \sum_{j=t+1}^{T} \|N_t(\mathbf{M}\mathbf{x}_j^d + \mathbf{w}) - \mathbf{F}_j^d\|$$
$$\mathbf{x}_{t+1}^{run} = \mathbf{M}_{t+1}\mathbf{x}_{t+1}^d + \mathbf{w}_{t+1}$$
$$(4)$$

Because $N_o(\cdot)$ is fully differentiable, and so is $N_t(\cdot)$, $\mathbf{M}$ and $\mathbf{w}$ can be solved suboptimally using gradient-based optimizers thanks to the differentiable neural representation of $N_o(\cdot)$. In the neckline enclosing setting, we found the simplification by setting $\mathbf{M} = \mathbf{I}$, i.e., restricting the same trajectory shape and only optimizing the translational shift yields more stable performance. Nevertheless, our formulation is generic and widely applicable to other tasks if the trajectory needs to undergo an affine transformation.

The complete steps of the neckline enclosing problem are summarized in Algorithm 2.

## IV. Experiments

### A. Implementation Details

We use a Kinova gen3 robotic arm platform equipped with a Robotiq 2F140 gripper. The right side of the hanger is inserted inside the T-shirt, and the end of the right T-shirt neckline is fixed on the hanger (Fig. 3). We collect the demonstration and the dataset $\mathbf{D}$ with only one T-shirt ((a) in Fig. 5) and the same grasp location.

An example of the grasp detection is shown in Fig. 4. In this case, $\delta$ is set to $0.2\%$ of the relative width of the gripper opening as the width variation factor to account for the uncertainties for the tactile image of the empty finger. From Fig., 4, we can see a tactile image with contact is distinctive from the one without contact, which helps to detect the contact. The primary source of difference is the textures of the cloth and the positional change of the markers due to the thickness of the cloth.

### D. Tactile Field Estimation

Because of varying grasp locations and changes in the manipuland, the position-force relationship from the collected data may differ at test time. Estimating the tactile field for the current configuration is desired to reproduce the force profile. To this end, we first train a nominal tactile field $\mathbf{F} = N_o(\mathbf{x})$ using the position-force dataset $\mathbf{D}$. We choose neural representation [22] for $N_o(\cdot)$ to have a universal and differentiable approximation. At test time $t$, we estimate the tactile field for the current configuration $N_t(\cdot)$ with measured $\{\mathbf{x}_j^{run}\}_{j=1}^t$ and $\{\mathbf{F}_j^{run}\}_{j=1}^t$ using the following.
**The tactile field optimization problem**:

$$\mathbf{A}_t, \mathbf{b}_t = \underset{\mathbf{A}, \mathbf{b}}{\operatorname{argmin}} \sum_{j=1}^{t} \|\mathbf{A}N_o(\mathbf{x}_j^{run}) + \mathbf{b} - \mathbf{F}_j^{run}\| \quad (3)$$
$$N_t(\cdot) = \mathbf{A}_t N_o(\cdot) + \mathbf{b}_t$$

**Algorithm 2:** The Enclosing Algorithm

---

**Input:** Demonstration trajectory $\{\mathbf{x}_i^d\}_{i=1}^T$,
demonstration force $\{\mathbf{F}_i^d\}_{i=1}^T$,
position-force dataset $\mathbf{D} = \{(\mathbf{x}_m^D, \mathbf{F}_m^D)\}_{m=1}^M$.
**Output:** A successful enclosure.

1 Train $N_o$ using $\mathbf{D}$;
2 Grasp the T-shirt using Algorithm 1;
3 $\mathbf{x}_1^{run} = \mathbf{x}_1^d$;
4 **for** $t = 1 : T$ **do**
5    Gripper goes to $\mathbf{x}_t^{run}$;
6    Measure $\mathbf{F}_t^{run}$;
7    **if** $t < 6$ **then**
8       $\mathbf{x}_{t+1}^{run} = \mathbf{x}_{t+1}^d$;
9    **else**
10       Estimate $N_t$ using Eq. 3;
11       **if** $t \neq T$ **then**
12          Determine $\mathbf{x}_{t+1}^{run}$ using Eq. 4;
13       **end**
14    **end**
15 **end**
16 **return**

---



Fig. 5: The different T-shirts used for the neckline enclosing problem. (a) A medium-sized T-shirt with a medium level of stiffness for collecting the demonstration and the dataset $\mathbf{D}$. (b) A small-sized soft T-shirt for testing. (c) A small-sized stiff T-shirt for testing. (d) A large-sized soft T-shirt for testing. (e) A large-sized stiff T-shirt for testing.

The demonstration is collected by a person to guide the robot arm to complete the neckline enclosing motion by setting the robot to the cartesian admittance mode. The positions of the gripper are recorded in a 1cm step size. Afterward, the 3D positions are projected into the 2D plane defined by the hanger. Then, the projected trajectory is executed again to ensure it is desirable and can accomplish the task. During this process, the shear force readings measured by the GelSight sensor are recorded at each corresponding waypoint.

To collect the dataset $\mathbf{D}$, a mesh grid is defined within the region of interest with 1cm sampling distance. The region of interest is coplanar with and upper bounded by the hanger. The collected region's lower boundary is 2cm below the grasp location. The region starts from the grasp location and extends leftward until the measured shear force reaches the limit of the tactile sensor. A graphical illustration of the region of interest for dataset $\mathbf{D}$ is shown as red in Fig. 1. A total of 581 sample points are collected in this process. We approximate the tactile field using a multilayer perceptron (MLP) with 3 hidden layers, each equipped with 256 neurons. To effectively utilize the dataset, we use all the collected data points to train the model and use linear interpolated data as a validation set to avoid overfitting. This particular operation is suitable in this context of the learned model because, eventually, we would want a differentiable model to interpolate the tactile field. 4 interpolated points between every other original point are generated as the validation dataset.

### B. Results

*1) Tactile Field Prediction:* The ability to estimate the current force field is evaluated with a designed examination trajectory. The examination trajectory is to firstly move hor-

izontally for 13cm from the grasp location, then diagonally towards the top left corner for 18cm. The force is measured along the trajectory at a step of every 1cm. When there are less than 6 points with a force magnitude greater than 0.2 in the history, the tactile field estimation is deactivated, and the original learned tactile field is returned to avoid uncertain estimation when there are not enough data points.

Fig. 6 shows that although the predicted tactile field deviates initially from the actual field due to differences between the actual T-shirt configuration/grasp location and the learned settings, our algorithm refines the prediction gradually as the trajectory is executed. The tactile field prediction also generalizes well to unseen T-shirts, as shown in Fig. 7. This might be due to:

1) Our grasp location identification method ensures the edge of the T-shirt is held regardless of its material, shape, and positional uncertainties, thus reducing the T-shirt configuration only to a subset.
2) The nonlinear elastic characteristics are encoded in the original learned tactile field, and as a result, a linear transformation of that field can closely approximate the new field very efficiently.

With the estimated tactile field as the elastic context of the task, we can then search for the proper trajectory to produce the desired force profile.

*2) Variation of Grasp Locations:* Fig. 8 shows the effect of varying the height of the search line for grasp by 2cm. When the grasp location changes, the nominal trajectory will keep the same position-based trajectory relative to the grasp location. We can see that the experimental results are consistent with the theoretical analysis in Sec. III-A. Specifically, a higher grasp location will generate a more significant force along the trajectory. This might damage the T-shirt or break the hanger. On the other hand, the force will be reduced for a trajectory with a lower grasp location, which may lead to insufficient stretching of the collar to accommodate the hanger (Fig. 3). In either case, our method (cross) reduces the uncertainty arising from variations in grasp location and tracks the demonstration force profile.
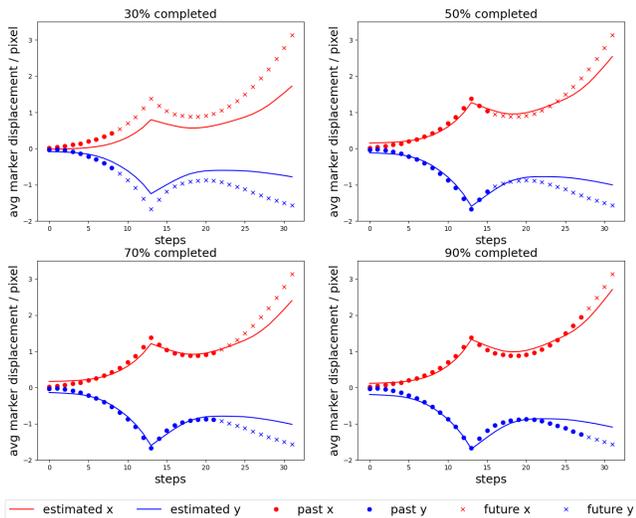
Fig. 6: The estimated and actual force at different stages on the examination trajectory for the learned T-shirt ((a) in Fig. 5).

*3) Comparison with Local Force Gradient Estimation:* We compared the proposed method with a baseline method that did not require the force-position dataset $\mathbf{D}$. The baseline method uses local exploration to estimate the force gradient and determine the direction and magnitude to adjust the gripper location to compensate for the difference between the current force measurement and the desired one from the demonstration. The exploration is achieved by moving around at the current location in two orthogonal axes in the 2D motion space. Each axis is sampled by a step size of 1cm in both positive and negative directions; therefore, 4 relative differences are used to estimate each local force gradient.

Fig. 9 compares the baseline and our proposed method with the same T-shirt and grasp location as the reference trajectory. There is minimal variation in the T-shirt configuration, implying that actual trajectory will not differ much from the demonstration trajectory when following the same force profile. Fig. 9 shows that the baseline method using local force gradient estimation is sensitive to noise in the tactile sensor, resulting in lower accuracy. Our method can estimate and predict the tactile field over the entire region. This enables planning over longer horizons and more stable behavior. Additionally, estimating the local gradient introduces extra motions at each step, which is not desirable in actual deployment.

*4) Tactile Trajectory Tracking:* To evaluate its generalization ability, we tested our proposed method with multiple T-shirts. Fig. 5 exhibits the T-shirts tested. T-shirts (b) and (d) are the softest and thinest ones that we can find. (c) and (e) are much thicker and stiffer than the learned T-shirt (a). The difference in the T-shirt stiffness can be quantitatively evaluated by looking at the force profile generated along the demonstration positions (nominal trajectory). The soft T-shirts generate a much smaller shear force. The peaks of the curves are less than 0.5 ((a) and (b) in Fig. 10). The stiff T-shirts produce much larger shear forces ((c) and (d)

in Fig. 10). The trajectory in Fig. 10(d) saturates because the maximum marker displacement of the sensor was exceeded. In all four cases in Fig. 10, our proposed method can adjust the trajectory to closely track the original demonstration force. This critical feature enables us to achieve the neckline enclosing motion and is potentially helpful for elastic deformable object robotic manipulations.

*5) Enclosure Task Success Rate:* Table. I summarizes the success rates of the neckline enclosure for different T-shirts in Fig. 5. We tested each T-shirt for 10 runs. The baseline fixed-position trajectory (nominal) failed many times for the stiff T-shirts, while with our proposed force tracking method, all of the tests succeeded. The main reason for the failure when following the fixed position trajectory is that the generated shear force often gets so large that the T-shirt slips out of the gripper. The large shear force is also potentially damaging to the hanger, the T-shirt, and the soft elastomer of the tactile sensor. Many trajectories with a successful enclosure potentially exist. Among them, our generated trajectory is simple to come up with, imitates a similar shape in space as the demonstration, and ensures a safe interaction with the T-shirt and the hanger. Our method tracks a moderate demonstration force to accomplish the task, making sure the force is regulated within a reasonable margin.

Table. I also tabulates the average distance between the hanger end and the collar end when the robot gripper is at the height of the hanger bottom, which we define as the safety margin. This measure quantifies the robustness against uncertainties. When the margin is large, the task is more likely to succeed for this T-shirt (because the hole is wide open) even if the configuration changes. For softer T-shirts, the proposed force-tracking method increases the safety margin significantly from the nominal trajectory because it pulls the T-shirt further away from the hanger. Although it seems less necessary in this experiment setting because the nominal trajectory with a smaller margin would still yield success, a larger margin gives extra space and robustness, especially when the fixed point (point O in Fig. 2) is further away to the hanger center. For stiffer T-shirts, the margin is reduced to decrease the generated shear force, but it does not vanish, leaving over 2cm space on average for the hanger to get in to account for other uncertainties in the setup.

TABLE I: Task execution statistics for the proposed method and baseline

| T-shirt type | Proposed | | Fixed Position | |
|---|---|---|---|---|
| | success rate | margin (cm) | success rate | margin (cm) |
| Trained T-shirt | 10/10 | 3.6 | 10/10 | 3.3 |
| Soft small | 10/10 | 5.9 | 10/10 | 4.2 |
| Soft large | 10/10 | 9.3 | 10/10 | 7.1 |
| Stiff small | 10/10 | 2.2 | 5/10 | 5.3 |
| Stiff large | 10/10 | 2.5 | 3/10 | 3.2 |

## V. CONCLUSION

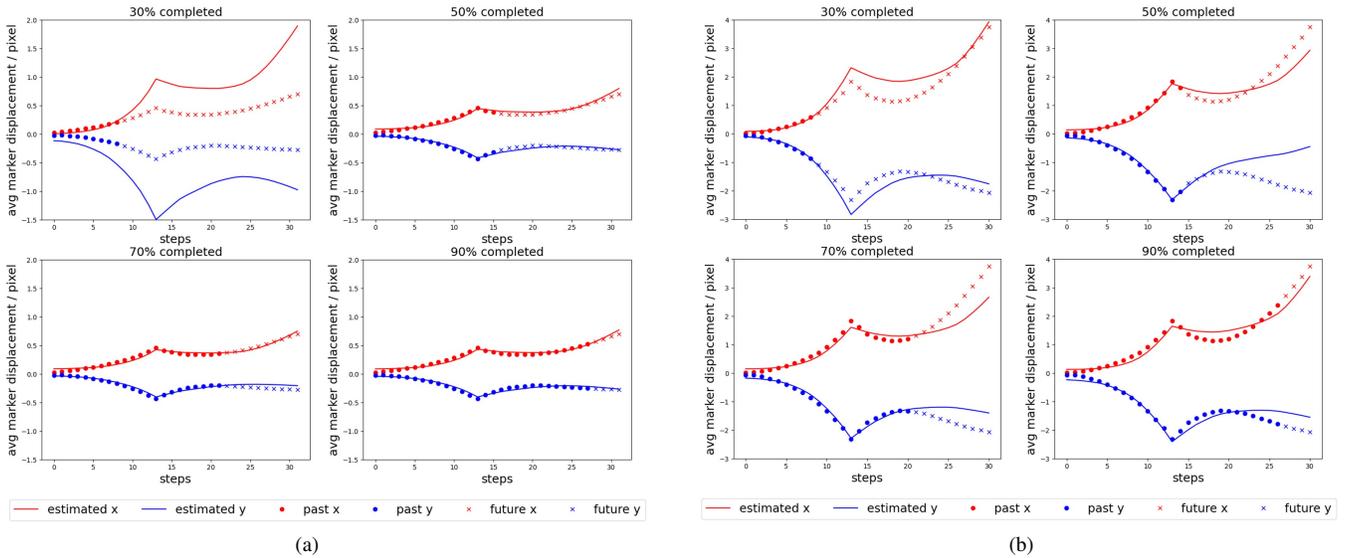In this work, we implemented a robotic system for deformable peg-in-hole manipulation with multimodal tactile

Fig. 7: The estimated and actual force at different stages on the examination trajectory for (a) the soft small T-shirt ((b) in Fig. 5) and (b) the small stiff T-shirt ((c) in Fig. 5).
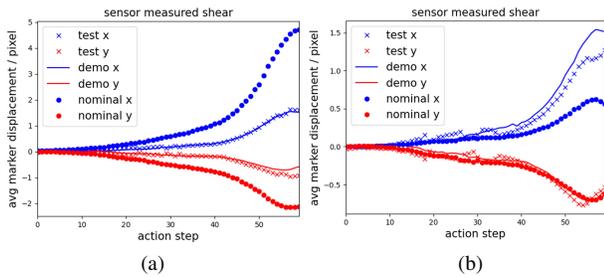


Fig. 8: (a): the force recorded when the search line for grasp is lifted by 2cm. (b): the force recorded when the search line for grasp is lowered by 2cm. Demo: the recorded force during the original demonstration (force to follow). Nominal: the force recorded if applying the demonstration trajectory. Cross: the force recorded on the optimized trajectory using the proposed method.
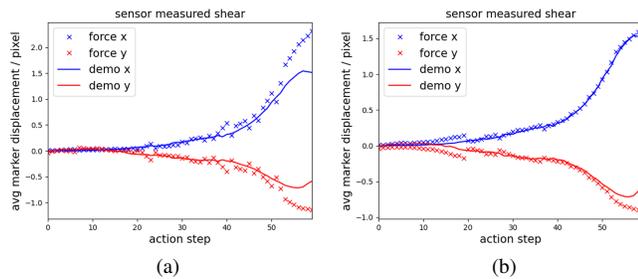


Fig. 9: Test to track the demonstration force with (a) local force gradient estimation and (b) our method.
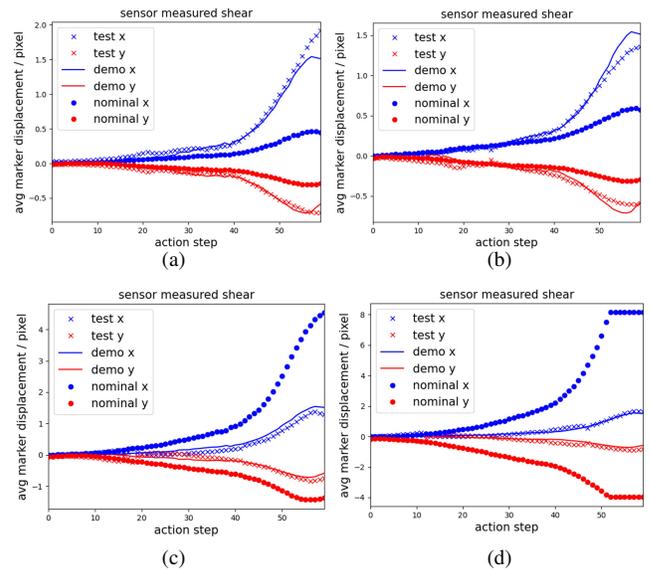


Fig. 10: Demo: the recorded force during the original demonstration (force to follow). Nominal: the force recorded if the demonstration trajectory is applied to the new T-shirt. Cross: the force recorded on the optimized trajectory using our proposed method. (a) With the soft, small T-shirt ((b) in Fig. 5). (b) With the soft large T-shirt ((d) in Fig. 5) (c) With the stiff small T-shirt ((c) in Fig. 5) (d) With the stiff large T-shirt ((e) in Fig. 5).

feedback and demonstrated it on the task of hanging T-shirts on hangers. We focused on two sub-problems: (1) how to find a proper grasp location on the T-shirt and (2) what trajectory to execute to accomplish the task. Different modalities of the tactile feedback of the GelSight sensor were utilized in two different stages. As a result, similar grasp locations (close to the cloth edge) and shear force trajectories were generated. The qualitative justifications for imitating

the force profile were illustrated. Among all the successful trajectories, our proposed one is easy to develop (with only one demonstration) and follows a blind human heuristic. The experiments demonstrated that our proposed strategy enables successful enclosure of the T-shirt collar around the hanger while keeping forces within a safe range and exhibiting robustness against configuration variations. One limitation is that we only compared our method with position-based trajectories and showed superiority in hanging the T-shirt.

However, our method work beyond hanging T-shirts; more generally, the force tracking algorithm is an approach to estimating and planning hybrid force-position trajectories when dealing with elastic and deformable objects, which we anticipate will apply to other deformable manipulation tasks. Overall, this work demonstrates that by combining the relative advantages of vision and tactile sensing, we can effectively reduce the complexity of deformable peg-in-hole manipulation, compared with using only visual feedback.

## ACKNOWLEDGMENT

## APPENDIX

### A. *Comparison of OA-OA' and OB-OB'*

Fig. 11 depicts the 2 triangles $\triangle OAA'$ and $\triangle OBB'$ formed by the trajectories $AB$ and $A'B'$. We can find a point $C$ on $OB$ such that $OC = OB'$, and similarly point C' so $OC' = OA'$. Connecting $B'C$ and $A'C'$, we can get $\triangle AA'C'$ and $\triangle BB'C$, in which

1) $AA' = BB'$
2) $\angle A'AC' < \angle B'BC$
3) $\angle A'OC' < \angle B'OC$(Given $OA \approx OB$)
$\Rightarrow \angle A'C'O = \frac{\pi - \angle A'OC'}{2} > \frac{\pi - \angle B'OC}{2} = \angle B'CO$
$\Rightarrow \angle AC'A' < \angle BCB'$

Given the above 3 conditions, we can plot the circumscribed circle of $\triangle AA'C'$ as in the right of Fig. 11. If we are to move point $C'$ to point $C$ to obtain $\triangle BB'C$, given condition 3), we can only move $C'$ to the inside of the circle. Moreover, to achieve condition 2), the choice of $C$ can only be in the red region. Thus, we must have $AC' > BC$. And therefore, $OA - OA' = AC' > BC = OB - OB'$
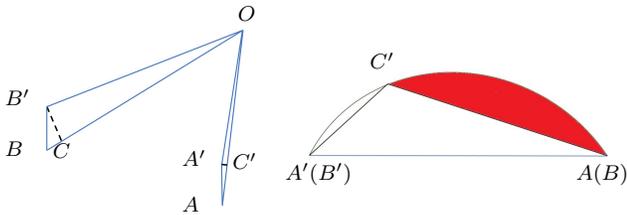


Fig. 11: Left: the 2 triangles from Fig. 2. Right: $\triangle AA'C'$ and its circumscribed arc (with $AA' = BB'$).

## REFERENCES

[1] H. Yin, A. Varava, and D. Kragic, "Modeling, learning, perception, and control methods for deformable object manipulation," *Science Robotics*, vol. 6, no. 54, p. eabd8803, 2021.

[2] J. Luo, O. Sushkov, R. Pevceviciute, W. Lian, C. Su, M. Vecerik, N. Ye, S. Schaal, and J. Scholz, "Robust Multi-Modal Policies for Industrial Assembly via Reinforcement Learning and Demonstrations: A Large-Scale Study," in *Proceedings of Robotics: Science and Systems (RSS)*, Virtual, July 2021.

[3] F. Zhang and Y. Demiris, "Visual-tactile learning of garment unfolding for robot-assisted dressing," *IEEE Robotics and Automation Letters*, 2023.

[4] S. Li, N. Figueroa, A. Shah, and J. Shah, "Provably safe and efficient motion planning with uncertain human dynamics," in *Proceedings of Robotics: Science and Systems (RSS)*, Virtual, July 2021.

[5] 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.

[6] L. Y. Chen, B. Shi, D. Seita, R. Cheng, T. Kollar, D. Held, and K. Goldberg, "Autobag: Learning to open plastic bags and insert objects," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 3918–3925.

[7] J. Sanchez, J. A. Corrales Ramon, B.-C. Bouzgarrou, and Y. Mezouar, "Robotic Manipulation and Sensing of Deformable Objects in Domestic and Industrial Applications: A Survey," *The International Journal of Robotics Research*, vol. 37, no. 7, pp. 688 – 716, Jun. 2018. [Online]. Available: https://uca.hal.science/hal-01816189

[8] J. Zhu, A. Cherubini, C. Dune, D. Navarro-Alarcon, F. Alambeigi, D. Berenson, F. Ficuciello, K. Harada, J. Kober, X. Li, J. Pan, W. Yuan, and M. Gienger, "Challenges and outlook in robotic manipulation of deformable objects," *IEEE Robotics & Automation Magazine*, vol. 29, no. 3, pp. 67–77, 2022.

[9] Y. She, S. Wang, S. Dong, N. Sunil, A. Rodriguez, and E. Adelson, "Cable manipulation with a tactile-reactive gripper," *The International Journal of Robotics Research*, vol. 40, no. 12-14, pp. 1385–1401, 2021. [Online]. Available: https://doi.org/10.1177/02783649211027233

[10] J. Zhu, B. Navarro, R. Passama, P. Fraisse, A. Crosnier, and A. Cherubini, "Robotic manipulation planning for shaping deformable linear objects withenvironmental contacts," *IEEE Robotics and Automation Letters*, vol. 5, no. 1, pp. 16–23, 2020.

[11] C. Chi, B. Burchfiel, E. Cousineau, S. Feng, and S. Song, "Iterative residual policy: For goal-conditioned dynamic manipulation of deformable objects," *The International Journal of Robotics Research*, vol. 0, no. 0, p. 02783649231201201, 0. [Online]. Available: https://doi.org/10.1177/02783649231201201

[12] K. Puthuveetil, C. C. Kemp, and Z. Erickson, "Bodies uncovered: Learning to manipulate real blankets around people via physics simulations," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 1984–1991, 2022.

[13] H. Ha and S. Song, "Flingbot: The unreasonable effectiveness of dynamic manipulation for cloth unfolding," in *Conference on Robotic Learning (CoRL)*, 2021.

[14] N. Sunil, S. Wang, Y. She, E. H. Adelson, and A. Rodriguez, "Visuotactile affordances for cloth manipulation with local control," in *Conference on Robot Learning*, 2022. [Online]. Available: https://api.semanticscholar.org/CorpusID:254564380

[15] W. Yuan, Y. Mo, S. Wang, and E. H. Adelson, "Active clothing material perception using tactile sensing and deep learning," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, 2018, pp. 4842–4849.

[16] F. Zhang and Y. Demiris, "Visual-tactile learning of garment unfolding for robot-assisted dressing," *IEEE Robotics and Automation Letters*, vol. 8, no. 9, pp. 5512–5519, 2023.

[17] M. Dragusanu, S. Marullo, M. Malvezzi, G. M. Achilli, M. C. Valigi, D. Prattichizzo, and G. Salvietti, "The dressgripper: A collaborative gripper with electromagnetic fingertips for dressing assistance," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 7479–7486, 2022.

[18] D. Leins, F. Patzelt, and R. Haschke, "More accurate tactile sensor simulation with hydroelastic contacts in mujoco," in *IROS 2023 Workshop on Leveraging Models for Contact-Rich Manipulation*, 2023.

[19] A. Longhini, M. Moletta, A. Reichlin, M. C. Welle, A. Kravberg, Y. Wang, D. Held, Z. Erickson, and D. Kragic, "Elastic context: Encoding elasticity for data-driven models of textiles," in *2023 IEEE International Conference on Robotics and Automation (ICRA 2023)*, 2023, pp. 1764–1770.

[20] R. Hanai, Y. Domae, I. G. Ramirez-Alpizar, B. Leme, and T. Ogata, "Force map: Learning to predict contact force distribution from vision," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2023, pp. 3129–3136.

[21] L. Y. Chen, H. Huang, E. Novoseller, D. Seita, J. Ichnowski, M. Laskey, R. Cheng, T. Kollar, and K. Goldberg, "Efficiently learning single-arm fling motions to smooth garments," in *Robotics Research*, A. Billard, T. Asfour, and O. Khatib, Eds. Cham: Springer Nature Switzerland, 2023, pp. 36–51.

[22] R. Gao, Y.-Y. Chang, S. Mall, L. Fei-Fei, and J. Wu, "Objectfolder: A dataset of objects with implicit visual, auditory, and tactile representations," in *5th Annual Conference on Robot Learning*, 2021.