

SIS: Seam-Informed Strategy for T-Shirt Unfolding

Xuzhao Huang ¹, Graduate Student Member, IEEE, Akira Seino ², Member, IEEE,
 Fuyuki Tokuda ³, Member, IEEE, Akinari Kobayashi ⁴, Member, IEEE,
 Dayuan Chen ⁵, Graduate Student Member, IEEE, Yasuhisa Hirata ⁶, Member, IEEE,
 Norman C. Tien ⁷, Senior Member, IEEE, and Kazuhiro Kosuge ⁸, Life Fellow, IEEE

Abstract—Seams are information-rich components of garments. The presence of different types of seams and their combinations helps to select grasping points for garment handling. In this letter, we propose a new Seam-Informed Strategy (SIS) for finding actions for handling a garment, such as grasping and unfolding a T-shirt. Candidates for a pair of grasping points for a dual-arm manipulator system are extracted using the proposed Seam Feature Extraction Method (SFEM). A pair of grasping points for the robot system is selected by the proposed Decision Matrix Iteration Method (DMIM). The decision matrix is first computed by multiple human demonstrations and updated by the robot execution results to improve the grasping and unfolding performance of the robot. Note that the proposed scheme is trained on real data without relying on simulation. Experimental results demonstrate the effectiveness and generalization ability of the proposed strategy.

Index Terms—Perception for grasping and manipulation, bimanual manipulation, manipulation planning.

I. INTRODUCTION

GARMENT unfolding remains an open challenge in robotics. Existing research utilizes folds [1], [2], [3], edges [4], [5], outline points, and structural regions [6], [7] as primary references for selecting grasping points, or uses a value map calculated using models trained in simulators [8], [9], [10], [11], [12]. However, none of these methods have utilized seam information. Seams are usually located in the contour position

Received 7 March 2025; accepted 12 May 2025. Date of publication 29 May 2025; date of current version 11 June 2025. This article was recommended for publication by Associate Editor S. Foix and Editor J. B. Sol upon evaluation of the reviewers' comments. This work was supported in part by the JC STEM Lab of Robotics for Soft Materials funded by The Hong Kong Jockey Club Charities Trust and in part by the Innovation and Technology Commission of the HKSAR Government through the InnoHK initiative. (Corresponding author: Xuzhao Huang.)

Xuzhao Huang, Akira Seino, Fuyuki Tokuda, Akinari Kobayashi, and Kazuhiro Kosuge are with the JC STEM Lab of Robotics for Soft Materials, Department of Electrical and Electronic Engineering, Faculty of Engineering, The University of Hong Kong, Hong Kong SAR, China, and also with the Centre for Transformative Garment Production, Hong Kong SAR, China (e-mail: x.z.huang@connect.hku.hk; kosuge@hku.hk).

Dayuan Chen and Yasuhisa Hirata are with the Centre for Transformative Garment Production, Hong Kong SAR 000000, China, and also with the Hirata Laboratory, Department of Robotics, Graduate School of Engineering, Tohoku University, Sendai 980-8579, Japan.

Norman C. Tien is with the Centre for Transformative Garment Production, Hong Kong SAR, China, and also with the Department of Electrical and Electronic Engineering, Faculty of Engineering, The University of Hong Kong, Hong Kong SAR 000000, China.

The project home page is available at <https://github.com/lancexz/sis>.

This article has supplementary downloadable material available at <https://doi.org/10.1109/LRA.2025.3574966>, provided by the authors.

Digital Object Identifier 10.1109/LRA.2025.3574966



Fig. 1. T-shirt unfolding strategy based on seam features constructed from seam line segments and their crossings. The candidates for a pair of grasping points for a robot motion primitive are limited to the seam features marked in the figure.

of the garment when fully unfolded. Consequently, the desirable grasping points for unfolding the garment tend to fall near the seams.

We believe that the introduction of seam information could improve the efficiency of the garment unfolding process for the following reasons:

- Seams can be used as a universal garment feature due to their prevalence in different types of garments.
- Seams are more visible than other features when the garment is randomly placed, facilitating the perception process without additional reconfiguration of the garment.
- The introduction of seam information makes it possible to select the grasping points without explicitly using the garment structure, resulting in efficient garment handling.

This letter presents Seam-Informed Strategy (SIS), a novel scheme for selecting actions that can be used for automatic garment handling, such as dual-arm robotic T-shirt unfolding. We use the seam segments and their crossings as reference information to limit the search space for selecting grasping points, as shown in Fig. 1.

To facilitate the SIS, we propose a Seam Feature Extraction Method (SFEM) and a Decision Matrix Iteration Method (DMIM). The SFEM is used to extract seam features as candidates for a pair of dual-arm grasping points. The DMIM is used as a comprehensive grasping points selection method from the extracted candidates for dual-arm unfolding motion planning. This motion involves a sequence of robot motion primitives including grasping, stretching, and flinging to flatten and align the T-shirt.

We train the networks in SFEM with real data annotated by humans. The decision matrix in DMIM is initially computed with human demonstrations and updated with real robot execution results. Using the decision matrix, the robot can also

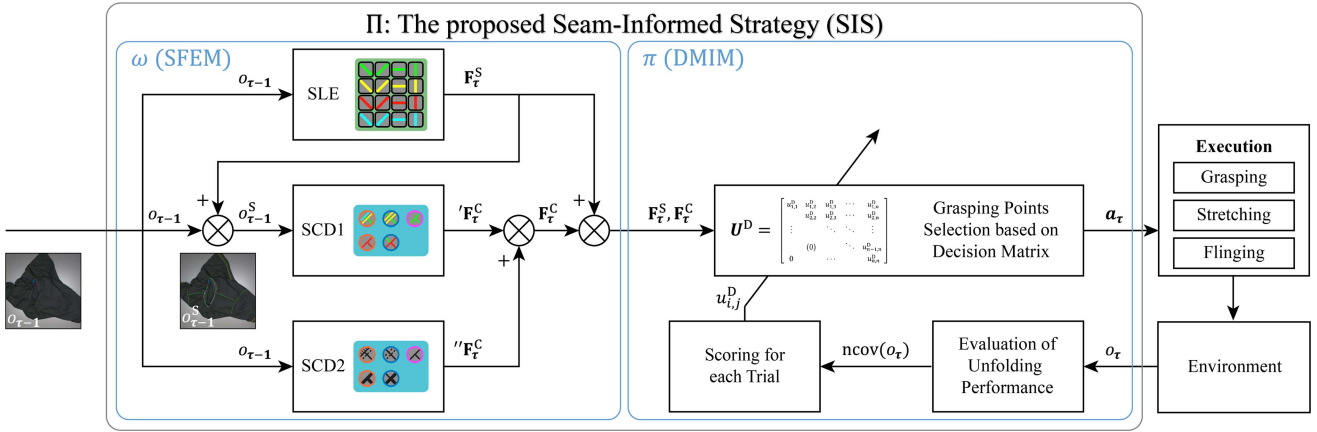


Fig. 2. The outline of the proposed SIS for unfolding a T-shirt from a randomly initialized configuration. ω and π are the methods introduced in III-B using the proposed Seam Feature Extraction Method (SFEM) and Decision Matrix Iteration Method (DMIM). Given an input of an original image observation $o_{\tau-1}$, SFEM, which consists of a Seam Line-segment Extractor (SLE) and two Seam Crossing-segment Detectors (SCD1 and SCD2), extract the seam line segment set F_{τ}^S and the seam crossing segment set F_{τ}^C based on YOLOv3 [21]. A seam-map $o_{\tau-1}^S$ is generated by superimposing the extracted F_{τ}^S onto $o_{\tau-1}$. DMIM selects the grasping points according to the scoring based on a decision matrix U^D , and outputs an a_{τ} , which is described by two grasping points $img_p_{\tau}^L$ and $img_p_{\tau}^R$. After performing the action a_{τ} with the motion primitives, grasping, stretching, and flinging, U^D is updated using the normalized coverage of the T-shirt $ncov(o_{\tau})$ in the observation o_{τ} .

align the orientation of the unfolded T-shirt. We evaluate the efficiency of our scheme on the real robot system through extensive experiments.

In summary, our contributions include:

- We propose SIS, a novel strategy for selecting grasping points for T-shirt unfolding using the seam information. SFEM and DMIM are proposed for this strategy.
- In the proposed SFEM, we formulate the extraction of seam lines as an oriented line object detection problem. This formulation allows the use of any object detection network to efficiently handle curved/straight seam lines.
- We solve the grasping points selection as a scoring problem for combinations of seam segment types using a proposed DMIM, a low-cost solution for unfolding a garment.
- Experimental results demonstrate that the performance of unfolding the T-shirt is promising in terms of obtaining high evaluation metrics with few episode steps.

II. RELATED WORK

Unfolding and flattening is the first process that enables various garment manipulation tasks. However, the selection of grasping points remains a challenging aspect of garment unfolding. Much work has been done so far.

A. Grasping Points Selection Strategies

There are three main types of strategies for selecting grasping points in garment unfolding or other garment handling tasks: heuristic-based strategy, matching-based strategy, and value-map-based strategy.

1) *Heuristic-Based Strategy*: A common way is to limit the possible garment configurations by first randomly picking up the garment and then grasping the lowest point [4], [13], [14]. A heuristic method proposed by [13] is to unfold a garment by repeatedly grasping the lowest point of the picked up garment.

2) *Matching-Based Strategy*: This strategy first maps the observed garment configuration to a canonical configuration and then obtains the grasping points based on the canonical

configuration [15]. The method proposed in [16] estimates the garment pose and the corresponding grasping point by searching a database constructed for each garment type.

3) *Value-Map-Based Strategy*: Recently, self-supervised learning frameworks [8], [9], [10], [11], [12], [17] have been used to predict action value maps or ranking scores that indicate the unfolding performance of grasping point candidates. [9], [10], [11] also consider the resulting garment orientation from the prioritized action.

B. Datasets for Learning-Based Methods

One of the key issues for learning-based methods is the lack of data sets when dealing with garments due to the difficulty of annotation. To address this problem, most researchers [8], [9], [10], [11], [12], [18] train their networks using simulators. SpeedFolding [17] uses real data annotated by humans together with self-supervised learning. In [4] and [6], the use of color-marked garments is proposed to automatically generate annotated depth maps. The use of transparent fluorescent paints and ultraviolet (UV) light is proposed by [19] to generate RGB images annotated by the paints observed under UV light.

C. Unfolding Actions Strategies

In addition to the commonly used Pick&Place, the Flinging proposed by FlingBot [8] is widely used for unfolding actions. FabricFolding [12] introduces a Pick&Drag action to improve the unfolding of sleeves. Meanwhile, DextAIRity [20] proposes the use of airflow provided by a blower to indirectly unfold the garment, thereby reducing the workspace of robots. Uni-Folding [11] uses a Drag&Mop action to reposition the garment when the grasping points are out of the robot's reach.

III. PROBLEM STATEMENT

This letter focuses on automating the process of unfolding a T-shirt from arbitrary configurations by a sequence of grasping, stretching, and flinging motions using a dual-arm manipulator system under the following assumptions:

- Seams are always present on the target T-shirt.
- The shirt is not in an inside-out configuration.

The desired outcomes include reducing the number of necessary steps of robot actions (episode steps) to unfold a T-shirt and increasing the normalized coverage of the final unfolded garment. We also consider the final orientation of the unfolded garment. This section describes the problem to be solved in this letter.

A. Problem Formulation

Given an observation $o_{\tau-1} \in \mathbb{R}^{W \times H \times 3}$, captured by an RGB camera with resolution of $W \times H$ at episode step $\tau \in \{1, 2, \dots\}$, our goal is to develop a policy Π mapping $o_{\tau-1}$ to an action a_τ , i.e. $a_\tau = \Pi(o_{\tau-1})$, to make the garment converge to a flattened configuration. The action a_τ can be described by the pair of grasping points for the left and right hands, $\text{img } \mathbf{p}_\tau^L \in \mathbb{R}^2$ and $\text{img } \mathbf{p}_\tau^R \in \mathbb{R}^2$, w.r.t the image frame Σ_{img} .

B. Outline of the Proposed Scheme

Existing strategies try to map the $o_{\tau-1}$ to a canonical space or a value map to select grasping points pixel-wise. This letter proposes a new Seam-Informed Strategy (SIS) to select grasping points based on seam information. To implement the mapping policy Π based on SIS, we divide the policy Π into two methods, ω and π , as shown in Fig. 2.

The method ω called the Seam Feature Extraction Method (SFEM) is proposed to extract seam line segments and their crossings from $o_{\tau-1}$. We use the extracted set of seam line segments \mathbf{F}_τ^S and the set of seam crossing segments \mathbf{F}_τ^C as candidates for grasping points. The method ω is formulated as follows:

$$\omega(o_{\tau-1}) = (\mathbf{F}_\tau^S, \mathbf{F}_\tau^C) \quad (1)$$

The method π called the Decision Matrix Iteration Method (DMIM) is proposed to select a pair of grasping points from the candidates obtained above for the unfolding action of the dual-arm robot. The decision matrix takes into account both human prior knowledge and the results of real robot executions. The method π is formulated as follows:

$$\pi(\mathbf{F}_\tau^S, \mathbf{F}_\tau^C) = (\text{img } \mathbf{p}_\tau^L, \text{img } \mathbf{p}_\tau^R) \quad (2)$$

From (1) and (2), we have:

$$\Pi(o_{\tau-1}) = \pi(\omega(o_{\tau-1})) = (\text{img } \mathbf{p}_\tau^L, \text{img } \mathbf{p}_\tau^R) \quad (3)$$

IV. METHOD

As outlined in Section III-B, this letter primarily addresses two challenges. In Section IV-A, we will explain how to extract sets of grasping point candidates \mathbf{F}_τ^S and \mathbf{F}_τ^C from an RGB image input $o_{\tau-1}$ using the proposed SFEM at each episode step τ . In Section IV-B, we will explain how to select a pair of grasping points $\text{img } \mathbf{p}_\tau^L$ and $\text{img } \mathbf{p}_\tau^R$ from the candidates \mathbf{F}_τ^S and \mathbf{F}_τ^C using the proposed DMIM at each episode step τ .

A. Seam Feature Extraction Method (SFEM)

The proposed SFEM consists of a seam line-segment extractor (SLE) and two seam crossing-segment detectors (SCD1 and SCD2) as shown in Fig. 2 which are designed based on the YOLOv3 [21]. Extracting seams as segments allows them to be used as grasping point candidates.

1) *Seam Line-Segment Extractor (SLE)*: The YOLOv3 was originally designed for object detection. For the SLE, we formulate the extraction of curved / straight seams as an oriented line object detection problem since the YOLOv3 cannot be used directly for seam line segment extraction. The proposed formulation allows any object detection network to extract seam line features.

To extract seams using the object detection network, we first approximate the continuous curved seam as a set of straight line segments. Each straight line segment \mathbf{f}^{SL} is described by the seam segment category $j \in \{1, 2, 3, 4\}$ and its endpoint positions (x_1, y_1) and (x_2, y_2) w.r.t the image frame Σ_{img} , i.e. $\mathbf{f}^{\text{SL}} = [j, x_1, y_1, x_2, y_2]$. In this letter, the seam segments are categorized into four categories, namely solid ($j = 1$), dotted ($j = 2$), inward ($j = 3$), and neckline ($j = 4$), as shown in Fig. 3(a) and (b), based on the seam types of the T-shirt used in the experiments.

We then convert the straight line segment \mathbf{f}^{SL} into a bounding box \mathbf{f}^{S} by using the two endpoints as the diagonal vertices of the bounding box. To represent the orientation of the straight line segment \mathbf{f}^{SL} in the bounding box \mathbf{f}^{S} , we divide each seam segment category into four orientation subclasses $i \in \{1, 2, 3, 4\}$, namely downward diagonal ($i = 1$), upward diagonal ($i = 2$), horizontal ($i = 3$), and vertical ($i = 4$), as shown in Fig. 3(a). This quadrupled the number of seam segment categories. Consequently, each oriented seam line segment $\mathbf{f}^{\text{S}} \in \mathbf{F}^{\text{S}}$ is defined by the oriented seam segment category $s_{i,j}$ and the parameters x, y, \hat{w}, \hat{h} of the bounding box, i.e. $\mathbf{f}^{\text{S}} = [s_{i,j}, x, y, \hat{w}, \hat{h}]$, where x, y denote the coordinate of the center of the bounding box. \hat{w}, \hat{h} denote the width and height of the bounding box.

This bounding-box-based categorization of the seam line segment is used as the labeling rule for the SLE shown in Algorithm 1. To implement the SLE, we introduce:

- λ_{thres} , a threshold for the width and height of the bounding box in pixels: Note that the computation of the loss function during network training is very sensitive to a tiny or thin object whose area is very small. λ_{thres} is introduced in Algorithm 1 to guarantee a minimum size of the bounding box.
- Data augmentation of \mathbf{f}^{S} with recategorization: The orientation subclass is not invariant if the image is flipped or rotated during data augmentation. Recategorization is carried out using Algorithm 1 when the image is flipped or rotated.

The proposed SLE can extract both curved and straight seams as a set of seam line segments. The predicted seam segments are superimposed onto the original observation $o_{\tau-1}$ to generate the seam-map $o_{\tau-1}^{\text{S}}$, as shown in Fig. 3(c)

2) *Seam Crossing-Segment Detectors (SCDs)*: As shown in Fig. 3(b), three types of seam crossing segments, namely shoulder (c_1), bottom hem (c_2), and neck point (c_3) of the T-shirt, are detected by using both SCD1 and SCD2 to increase the recall of the predictions. The inputs of SCD1 and SCD2 are the original image observation $o_{\tau-1}$ and the corresponding seam-map $o_{\tau-1}^{\text{S}}$, respectively, as shown in Fig. 2.

Note that the outputs of SCD1 and SCD2 are merged considering the maximum number of each type of crossing segments according to the confidence of the predictions, since the number of each type of seam crossing segments on the T-shirt is limited.

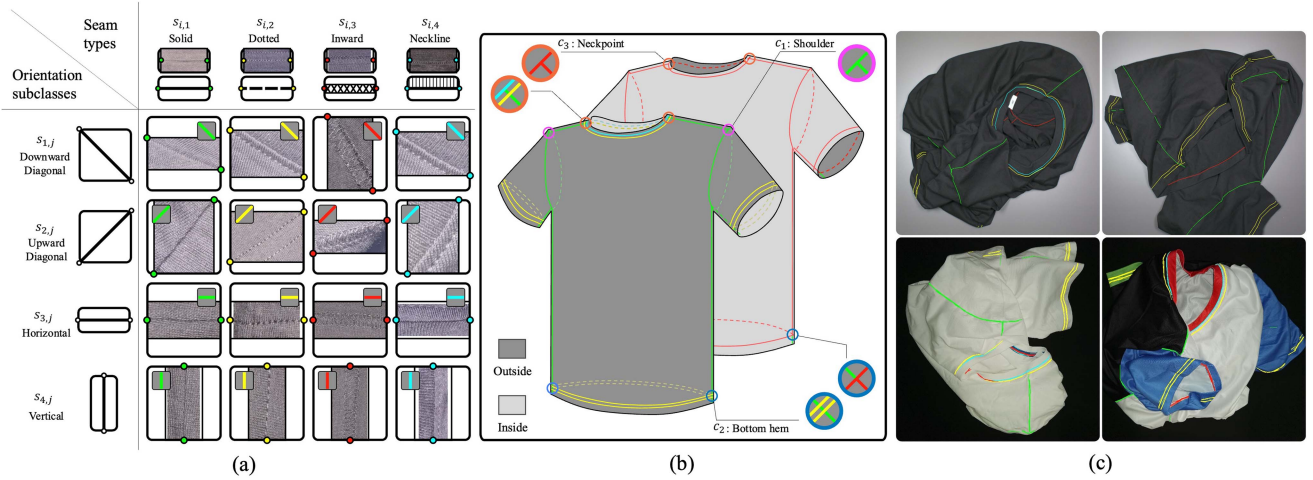


Fig. 3. Categorization of the grasping point candidates in the proposed SFEM. The four different colored lines represent four different types of seam segments: solid, dotted, inward, and neckline. (a) The categorization method for orientated seam line objects. To extract seam segments with orientation information, $s_{i,j}$ is used to denote a two-dimensional categorization for SLE in Algorithm 1. After dividing the seams into four types (solid, dotted, inward, and neckline), we further divide each type into four subclasses based on orientation (downward diagonal, upward diagonal, horizontal, and vertical). (b) The distribution of each type of seam segment and the definition of their crossings in canonical space. In this figure, solid lines denote visible seams, while dashed lines denote occluded seams. Note that the seam structure shown here is only an example. T-shirts with different seam structures are included in the training dataset. (c) Visualization examples of SFEM output on real T-shirts. The visualized images are named seam-maps in this letter.

Algorithm 1: Bounding-box-based categorization of SLE

Require: $f^{\text{SL}} = [j, x_1, y_1, x_2, y_2], \lambda_{\text{thres}}$
Ensure: $f^{\text{S}} = [s_{i,j}, x, y, \hat{w}, \hat{h}]$
 $x, y, w, h = \text{convert_to_boundingbox}(x_1, y_1, x_2, y_2)$
 $\hat{w} = w, \hat{h} = h$
if ($w < \lambda_{\text{thres}} \ \& \ h < \lambda_{\text{thres}}$):
 $\text{drop_this_label}()$
elif ($w < \lambda_{\text{thres}} \ \& \ h > \lambda_{\text{thres}}$):
 $i = 4, \hat{w} = \text{int}(h/2)$ ▷ Vertical
elif ($w > \lambda_{\text{thres}} \ \& \ h < \lambda_{\text{thres}}$):
 $i = 3, \hat{h} = \text{int}(w/2)$ ▷ Horizontal
elif ($x_1 < x_2 \ \& \ y_1 > y_2$) **or** ($x_1 > x_2 \ \& \ y_1 < y_2$):
 $i = 2$ ▷ Upward Diagonal
else :
 $i = 1$ ▷ Downward Diagonal
 $s_{i,j} = \text{get_orientated_category}(i, j)$
return $[s_{i,j}, x, y, \hat{w}, \hat{h}]$

B. Decision Matrix Iteration Method (DMIM)

The idea of DMIM is proposed based on the observations of human unfolding actions using a flinging motion:

- The Combination of Seam Segment Types (CSST) of two selected grasping points affects the unfolding performance. For example, grasping two points both at the shoulders or bottom hems results in a higher unfolding quality than the other CSSTs.
- The current step of the unfolding action affects the unfolding performance of subsequent actions, since an appropriate action simplifies the complexity of the subsequent steps. To obtain a simpler configuration for subsequent actions, we intuitively select two grasping points that are empirically far apart.

Based on these observations, we extract the human skill of unfolding by scoring the performance of each CSST through

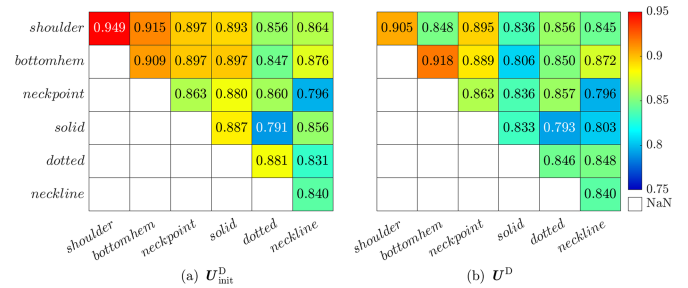


Fig. 4. The decision matrices used in the experiments. The matrices represent the unfolding performance score for each Combination of Seam Segment Type (CSST). (a) The initial decision matrix $U^{\text{D}}_{\text{init}}$ is computed with human demonstration data. (b) The updated decision matrix U^{D} is updated with data from robot trials.

human demonstrations. The DMIM proposed in this letter represents the grasping strategy using a decision matrix. The performance score of each CSST is implemented by initializing the decision matrix with human demonstrations and updating it with robot executions.

1) *Decision Matrix:* As shown in Fig. 4, the proposed decision matrix U^{D} is constructed as an upper triangular matrix. In this letter, six types of seam segments are considered, including shoulder, bottom hem, neck point, solid, dotted, and neckline (the inward seam line type is treated as the dotted type in this letter for simplicity). The (k, l) element of U^{D} , $u^{\text{D}}_{k,l}$, is the performance score of the combination of the k th and the l th seam segment types, and expressed as follows:

$$u^{\text{D}}_{k,l} = \frac{1}{M_{k,l}} \sum_{d=1}^{M_{k,l}} R_{d,k,l} \quad (4)$$

where $M_{k,l}$ is the total number of trials of unfolding demonstrations performed by both human and robot. Note that $k, l \in \{1, 2, \dots, 6\}$ and $k \geq l$. For each trial, the average normalized

coverage $R_{d,k,l}$ is expressed as follows:

$$R_{d,k,l} = \frac{1}{T_{d,k,l}} \sum_{\tau=1}^{T_{d,k,l}} \text{ncov}(o_{\tau}) \quad (5)$$

where $T_{d,k,l}$ is the number of episode steps in a trial and $\text{ncov}(o_{\tau})$ is the normalized coverage of the T-shirt from an overhead camera observation o_{τ} at episode step τ during a trial, similar to [8], [17], and [20]. The normalized coverage is calculated by:

$$\text{ncov}(o_{\tau}) = \frac{\text{cov}(o_{\tau})}{\text{cov}_{\max}} \quad (6)$$

where the coverage $\text{cov}(\cdot)$ is calculated by counting the pixel number of segmented mask output by the Segment Anything Model (SAM) [22]. The maximum coverage cov_{\max} is obtained from an image of a manually flattened T-shirt.

The decision matrix U^D is used to select grasping points from the seam segments extracted by SFEM. Based on the score of $u_{k,l}^D$ in the decision matrix and the types of extracted seam segments, the CSST with the highest score is selected. If there are multiple combinations of candidates belonging to each seam segment type of the selected CSST, the most distant candidate pair is selected as the grasping points.

2) *Initialization With Human Demonstrations and Updating With Robot Executions*: The initial decision matrix U_{init}^D , shown in Fig. 4(a), is computed using only the human demonstration data by (4). The same equation is then used to update the decision matrix with subsequent robot executions, resulting in U^D , as shown in Fig. 4(b).

Human hands exhibit more dynamic and flexible movements that allow fine-tuning of the resulting garment configuration, a capability that the robot manipulators lack. In other words, there is a human-to-robot gap in garment manipulation. To bridge this gap, we perform real robot trials to update the decision matrix. This iterative method ensures that the grasping strategy integrates the learned human experience and the real robot explorations.

V. EXPERIMENT

The garment unfolding dual-arm manipulator system consists of two Denso VS087 manipulators mounted on a fixed frame, a Basler acA4096-30uc RGB camera equipped with an 8 mm Basler lens, an Azure Kinect DK depth camera, two ATI Axia80-M8 Force Torque (F / T) sensors, and two Taiyo EGS2-LS-4230 electric grippers. The robot control system is constructed based on ROS2 Humble [23] installed on a PC equipped with an RTX3090 GPU, an i9-10900KF CPU, and 64 GB of memory.

A. Motion Primitives Used for Experiments

Below are the major motion primitives associated with the T-shirt unfolding experiments.

1) *Grasp&fling Motion Primitive*: Similar to [8], we implement a Grasp&Fling motion primitive to speed up the process of unfolding the garment. After grasping, the T-shirt is stretched by the robot arms until the stretching force reaches a predefined 0.8 N using the wrist F / T sensors attached to both manipulators. The robot then generates a flinging motion that mimics the human flinging motion based on fifth-order polynomial interpolation.

2) *Randomize Motion Primitive*: To generate initial configurations of the T-shirt that are fair enough for comparison, we

randomly select a grasping point from the garment area and release it from a fixed position 0.78 m above the desk using one of the robot arms. During the experiments described in Section V-C and V-D, the robot repeats this motion before each trial until the normalized coverage of the initial configuration is less than 0.4.

B. Training/Updating of SFEM and DMIM

1) *Training of SFEM*: The dataset used to train the SLE consists of 328 images with manually annotated labels of seam line segments. Using data augmentation by image rotation and flipping, we obtain 2104 images for training and 520 images for validation with annotated labels. We evaluated the seam line extraction performance of the trained model using a testing set in which garments were randomly placed on the desk. We calculated the Intersection over Union (IoU) between the predicted seam line segments and the manually labeled ground truth, using a line thickness of 20 pixels. The average IoU for seam line extraction is 0.734. Considering the nature of thin objects, this IoU result is reasonable as visualized in Fig. 3(c).

The dataset used to train SCD1 consists of 1708 original images with manually annotated labels of seam crossing segments. The dataset used to train SCD2 consists of the same 1708 images overlaid with the seam lines extracted by the trained SLE. We evaluated the seam crossing detection performance of the merged outputs of trained SCDs using precision and recall. The average precision is 0.915 and the average recall is 0.878 during our experiments.

The original images in all datasets are captured by the Basler camera. The resolution of the images used for training is 1280×1280 , obtained by cropping and resizing the original images.

2) *Initialization and Updating of DMIM*: The decision matrix U_{init}^D is initialized with human demonstrations. For each human trial, we set the maximum number of episode steps to $T_{d,k,l} = 5$, although two to three steps are usually enough to flatten the T-shirt. Once the T-shirt is flattened, we skip the remaining steps and use the same $\text{ncov}(\cdot)$ as in this step for the remaining steps.

To initialize a $u_{k,l}^D$ for each CSST, the demonstrator is asked to randomize the T-shirt until at least one point pair in the CSST appears. The demonstrator then selects the furthest pair of points as grasping points. In the following steps of this trial, the demonstrator selects grasping points based on the demonstrator's intuition, and the $u_{k,l}^D$ of the CSST is calculated. The matrix U_{init}^D is initialized as shown in Fig. 4(a). Each $u_{k,l}^D$ is computed using (4) with the number of trials $M_{k,l} = 10$.

To update the decision matrix, we conduct robot trials to bridge the human-to-robot gap in unfolding actions. Each robot trial consists of five episode steps. As shown in Fig. 4(b), U^D is updated from U_{init}^D with (4).

C. Ablation Studies of the Proposed Strategy

In this section, we perform ablation studies for our proposed strategy. In the following experiments, 20 trials are performed using two trained clothes in Category T in Fig. 7, with each trial consisting of five episode steps. The results are shown in Table I and Fig. 5. We use normalized coverage [8], [17], [20] and IoU between the current configuration and the manually unfolded configuration [9] as evaluation metrics. Fig. 5 shows the distributions of normalized coverage with 95% confidence interval shading.

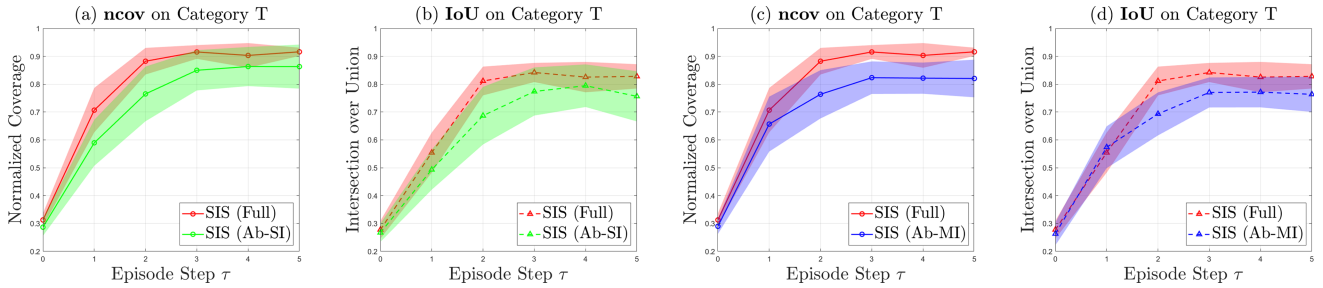


Fig. 5. Results of ablation studies using the methods proposed in this letter. The settings for each group in the plots can be found in Table I. We plot the normalized coverage in (a) and (c). The IoU between the current configuration and the goal configuration are plotted in (b) and (d). The curves in the plots denote the mean values of the metric, while the shaded areas represent the 95% confidence intervals describing the distribution of the experimental results. The data of the Full group represent the performance of our entire proposed method, while the other groups represent the performance after removing a specific module. Each group undergoes a total of 20 trials of experiments using T-shirts T-1 and T-2 in Fig. 7, where each trial contains five episode steps.

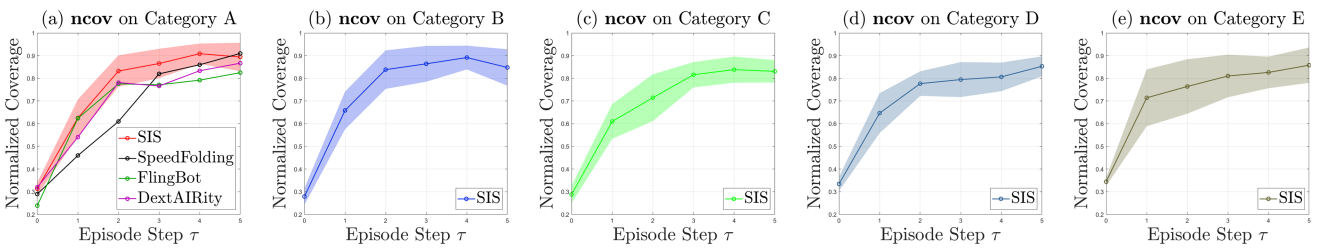


Fig. 6. Unfolding performance evaluation on normalized coverage using Categories A, B, C, D, and E garments in Fig. 7 and comparison with existing research using Category A. The data we use for comparison are those presented in their original papers of FlingBot [8], DextAIRity [20], and SpeedFolding [17].



Fig. 7. Clothes we used in the experiments. Category T is a set of 20 short-sleeved T-shirts used to train the network. T-1 and T-2 are two examples in Category T. Category A contains four unseen short-sleeved T-shirts that are not part of the training dataset. Categories B, C, D, and E contain unseen types of garments. Category B contains short-sleeved T-shirts with a sewn pocket. Category C includes two short-sleeved sports T-shirts that do not have all their seams in the contour area. Category D includes long-sleeved garments. Category E includes tank tops. Seams are highlighted with colored lines.

In Table I, the Full group is the complete scheme proposed in this letter based on seam information and DMIM using U^D .

The Ab-SI group is carried out without the seam information extracted by the SLE and uses only the crossings detected by the SCD2 as candidate grasping points, while U^D is used. The difference between the Ab-SI and Full groups in Fig. 5(a) and (b) shows that the seam information significantly improves the unfolding performance.

The Ab-MI group is performed to show the effectiveness of the proposed iteration method. In this experiment, U_{init}^D is used without updating. Fig. 5(c) and (d) shows the human-to-robot gap in the performance of the unfolding action. The results of the

Full and Ab-MI groups show that our proposed DMIM bridges the human-to-robot gap.

We conduct a right-tailed Welch's t-test to analyze the differences in the means of the ncov and IoU metrics before and after removing one component from the full scheme. Each group consists of 20 samples at each episode step. The null hypothesis H_0 posits that the Full group does not outperform the ablated configuration. The alternative hypothesis H_1 posits that the Full group significantly outperforms the ablated configuration. We reject the null hypothesis H_0 in favor of the alternative hypothesis H_1 when the p-value p is less than the significance level $\alpha = 0.05$, concluding that the performance of the Full

TABLE I
EVALUATION METRICS OF THE ABLATION EXPERIMENTS

Exp. Name	Seam Information (SI)	Matrix Iteration (MI)	Average evaluation metric at each step over 20 trials (ncov, IoU)				
			1	2	3	4	5
SIS (Full)	✓	✓	0.707 , 0.554	0.882 , 0.811	0.916 , 0.842	0.903 , 0.825	0.916 , 0.828
SIS (Ab-SI)		✓	0.590, 0.493	0.765, 0.687	0.850, 0.774	0.863, 0.795	0.863, 0.756
SIS (Ab-MI)	✓		0.657, 0.574	0.764, 0.693	0.823, 0.770	0.822, 0.771	0.820, 0.764

TABLE II
STATISTICAL ANALYSIS RESULTS OF P-VALUES USING WELCH'S T-TEST

Configuration	Episode Step				
	1	2	3	4	5
Ab-SI vs. Full					
ncov	0.026*	0.022*	0.053	0.177	0.106
IoU	0.123	0.022*	0.082	0.262	0.086
Ab-MI vs. Full					
ncov	0.222	0.013*	0.004**	0.016*	0.007**
IoU	0.642	0.009**	0.018*	0.089	0.055

* Statistically significant performance decrease ($p < 0.05$)

** Statistically highly significant performance decrease ($p < 0.01$)

group is significantly superior to that of the Ab-SI or Ab-MI group.

As shown in Table II, removing the Seam Information (SI) component results in statistically significant decreases in ncov during the early steps ($p < 0.05$). Similarly, removing the Matrix Iteration (MI) component leads to statistically significant decreases in ncov in later steps, with some decreases being highly significant ($p < 0.01$). The IoU results exhibit similar trends. Notably, both groups show significant performance drops during the second step, highlighting the advantages of our proposed method in addressing challenging scenarios, particularly when the garment is in a crumpled state.

Note that the following hardware-related failures are excluded from the statistical data since they are not relevant to the performance of our proposed strategy.

- Grasping failures: We consider a grasp to have failed if either one or both of the grippers failed to grasp the selected grasping points on the T-shirt. We also consider the grasp to have failed if the T-shirt falls off the gripper(s) during the robot motion. Note that similar to [8], we have not filtered out the data when the T-shirt is grasped in its crumpled state, which makes the flinging motion ineffective.
- Motion failures: These occur when the grasping points exceed the working space of the robot, or when an emergency stop is triggered due to robot singularity.
- Releasing failures: We consider a garment release to have failed if the garment remains in the gripper after the gripper fingers open.

D. Evaluation of Unfolding Performance

To evaluate the generalization ability of our proposed scheme, we conducted experiments with four unseen short-sleeved T-shirts in Category A in Fig. 7. In Fig. 6(a), we plotted the average result of normalized coverage in 20 unfolding trials, each with five episode steps. Five continuous unfolding trials are performed for each of the four T-shirts. On average, the normalized coverage reaches over 0.865, 0.909, and 0.894 within three, four, and five steps, respectively.

We compare the proposed method in this letter with those presented in [8] [17][20], utilizing normalized coverage as the

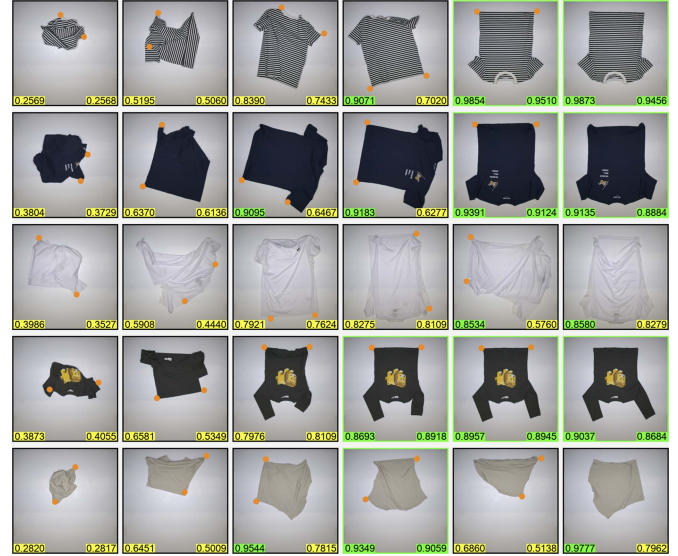


Fig. 8. Examples of experimental results. Due to space limitations, for each unseen category we show only the first of its continuous trials, each consisting of five episode steps. The orange dots indicate the selected grasping points for the dual-arm robot. For each image observation, the number in the lower left corner shows the normalized coverage of the T-shirt, while the number in the lower right corner shows the IoU result. Numbers are colored green if the metric has exceeded a threshold of 0.85. If both metrics exceed the threshold, the step is marked with a green border.

evaluation metric. As shown in Fig. 6(a), our SIS scheme outperforms existing methods. Note that our approach is simpler by using only a Grasp&Fling motion, while SpeedFolding needs additional motions such as Pick&Place and Drag in later steps. In addition, the proposed SIS aligns the T-shirt to a specific goal configuration, which SpeedFolding does not consider. We use IoU of resulting configuration with goal configuration to evaluate the orientation alignment performance on Category A. Our proposed SIS reaches 0.799, 0.826, and 0.831 average IoU within three, four, and five episode steps, respectively.

In Fig. 6(b) and (c), we have also plotted the normalized coverage of Category B and C. Category B contains two short-sleeved T-shirts with a pocket attached to the front with seams. Category C contains two short-sleeved sports T-shirts that have solid seams that are not in the contour area when fully unfolded. According to Welch's t-test, there is no significant difference in performance between experiments using Categories A and B. Therefore, we conclude that the seams on the pocket do not significantly affect the unfolding performance of our scheme. This is because the priority of selecting the dotted seams corresponding to the pockets is low, as shown in Fig. 4. Even for the challenging cases of Category C, the average coverage in ten trials of our scheme reaches more than 0.83 within five episode steps, demonstrating that our scheme can also be applied to such types of sports T-shirts.

Using the same models without additional training/updates, we conduct experiments on different types of garments, such as long-sleeved T-shirt and sleeveless shirt, shown in Category D and E in Fig. 7. A total of ten unfolding trials are performed on two cloth samples for each type of garment. The results are shown in Fig. 6(d) and (e). Although the performance suffers some degradation, our scheme can still unfold the garments without any additional retraining/updates. We conclude that our scheme can be applied to these two types of garments. Example results of unfolding trials are shown in Fig. 8.

VI. CONCLUSION

In this letter, we propose a Seam-Informed Strategy (SIS), which uses seam information to select a pair of grasping points for unfolding a T-shirt with a dual-arm robot system. Our strategy uses SFEM to extract grasping point candidates using seam information and DMIM to select the grasping points for unfolding the T-shirt while aligning its orientation in a low-cost manner. Experimental results using various unseen T-shirts and unseen types of garments have shown that the proposed SIS effectively handles T-shirt unfolding. The performance of T-shirt unfolding is promising in terms of obtaining high evaluation metrics with few episode steps.

Our future work aims to address the limitations of the current scheme when handling seamless garments, such as knitted garments, to improve the practicality and efficiency of our systems.

REFERENCES

- [1] D. Triantafyllou, I. Mariolis, A. Kargakos, S. Malassiotis, and N. Aspragathos, "A geometric approach to robotic unfolding of garments," *Robot. Auton. Syst.*, vol. 75, pp. 233–243, 2016.
- [2] A. Doumanoglou et al., "Folding clothes autonomously: A complete pipeline," *IEEE Trans. Robot.*, vol. 32, no. 6, pp. 1461–1478, Jun. 2016.
- [3] J. Stria, V. Petřík, and V. Hlaváč, "Model-free approach to garments unfolding based on detection of folded layers," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2017, pp. 3274–3280.
- [4] A. Gabas and Y. Kita, "Physical edge detection in clothing items for robotic manipulation," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2017, pp. 524–529.
- [5] X. Lin, Y. Wang, Z. Huang, and D. Held, "Learning visible connectivity dynamics for cloth smoothing," in *Proc. Mach. Learn. Res.*, vol. 164, pp. 08–11, Nov. 2022, pp. 256–266.
- [6] W. Chen, D. Lee, D. Chappell, and N. Rojas, "Learning to grasp clothing structural regions for garment manipulation tasks," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2023, pp. 4889–4895.
- [7] R. Wu, H. Lu, Y. Wang, Y. Wang, and H. Dong, "UniGarmentManip: A unified framework for category-level garment manipulation via dense visual correspondence," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2024, pp. 16340–16350.
- [8] H. Ha and S. Song, "FlingBot: The unreasonable effectiveness of dynamic manipulation for cloth unfolding," in *Proc. Conf. Robot Learn.*, 2021, pp. 24–33.
- [9] A. Canberk et al., "Cloth funnels: Canonicalized-alignment for multi-purpose garment manipulation," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2023, pp. 5872–5879.
- [10] N. Gu, R. He, and L. Yu, "Learning to unfold garment effectively into oriented direction," *IEEE Robot. Autom. Lett.*, vol. 9, no. 2, pp. 1051–1058, Feb. 2024.
- [11] H. Xue, Y. Li, W. Xu, H. Li, D. Zheng, and C. Lu, "Unifolding: Towards sample-efficient, scalable, and generalizable robotic garment folding," in *Proc. Conf. Robot. Learn.*, 2023, pp. 3321–3341.
- [12] C. He, L. Meng, Z. Sun, J. Wang, and M. Q.-H. Meng, "Fabricfolding: Learning efficient fabric folding without expert demonstrations," *Robotica*, vol. 42, no. 4, pp. 1281–1296, 2024.
- [13] F. Osawa, H. Seki, and Y. Kamiya, "Unfolding of massive laundry and classification types by dual manipulator," *J. Adv. Comput. Intell. Intell. Inform.*, vol. 11, no. 5, pp. 457–463, 2007.
- [14] A. Doumanoglou, T.-K. Kim, X. Zhao, and S. Malassiotis, "Active random forests: An application to autonomous unfolding of clothes," in *Proc. Eur. Conf. Comput. Vis.*, Springer, 2014, pp. 644–658.
- [15] C. Chi and S. Song, "Garmentnets: Category-level pose estimation for garments via canonical space shape completion," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 3324–3333.
- [16] Y. Li et al., "Regrasping and unfolding of garments using predictive thin shell modeling," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2015, pp. 1382–1388.
- [17] Y. Avigal, L. Berscheid, T. Asfour, T. Kröger, and K. Goldberg, "Speed-folding: Learning efficient bimanual folding of garments," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2022, pp. 1–8.
- [18] D. Tanaka, S. Arnold, and K. Yamazaki, "EMD Net: An encode-manipulate-decode network for cloth manipulation," *IEEE Robot. Autom. Lett.*, vol. 3, no. 3, pp. 1771–1778, Jul. 2018.
- [19] L. Y. Chen et al., "Autobag: Learning to open plastic bags and insert objects," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2023, pp. 3918–3925.
- [20] Z. Xu, C. Chi, B. Burchfiel, E. Cousineau, S. Feng, and S. Song, "Dextairity: Deformable manipulation can be a breeze," in *Proc. Robot. Sci. Syst.*, New York City, NY, USA, Jun. 2022, doi: [10.15607/RSS.2022.XVIII.017](https://doi.org/10.15607/RSS.2022.XVIII.017).
- [21] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," Apr. 2018, *arXiv:1804.02767*.
- [22] A. Kirillov et al., "Segment anything," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Oct. 2023, pp. 4015–4026.
- [23] S. Macenski, T. Foote, B. Gerkey, C. Lalancette, and W. Woodall, "Robot operating system 2: Design, architecture, and uses in the wild," *Sci. Robot.*, vol. 7, no. 66, 2022, Art. no. eabm6074.