Cloth Region Segmentation for Robust Grasp Selection

Jianing Qian*, Thomas Weng*, Brian Okorn, Luxin Zhang, and David Held
Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15213
{jianingq, tweng, bokorn, luxinz, dheld}@andrew.cmu.edu

***Abstract*—Cloth detection and manipulation is a common task in domestic and industrial settings, yet such tasks remain a challenge for robots due to cloth deformability. Furthermore, in many cloth-related tasks like laundry folding and bed making, it is crucial to manipulate specific regions like edges and corners, as opposed to folds. In this work, we focus on the problem of segmenting and grasping these key regions. Our approach trains a network to segment the edges and corners of a cloth from a depth image, distinguishing such regions from wrinkles or folds. We also provide a novel algorithm for estimating the grasp location, direction, and directional uncertainty from the segmentation. We demonstrate our method on a real robot system and show that it outperforms baseline methods on grasping success. Video and other supplementary materials are available at: https://sites.google.com/view/cloth-segmentation.**

I. Introduction

In cloth manipulation tasks such as laundry folding, it is important that the robot can identify and grasp key regions of the cloth. These regions typically include the "real edges" or corners of a cloth. By "real edges," we mean the edges of the cloth in the unfolded configuration, as opposed to any folds or creases that may appear as edges in a particular configuration. If the robot grasps a cloth fold or crease and attempts to use such a grasp to neatly fold the cloth, the result likely will not end up as expected. Thus, failing to grasp the cloth along the real edges could lead to failures for many downstream tasks.

As we will show, traditional computer vision algorithms fail to distinguish the difference between real cloth edges and apparent edges created by creases or folds. In addition, the robot must also determine the appropriate grasping direction along the cloth edge, which is non-trivial if the cloth is in a crumpled configuration; we will show that simple heuristics frequently fail at this task. We provide a method that identifies edges and corners of a cloth, predicts grasp directions, and estimates the uncertainty of these directions. These predictions will then be used to quickly and reliably grasp the cloth along its edges and corners, even from crumpled configurations.

In this paper, we present an approach for segmenting these key regions of cloth, even in highly crumpled configurations (see Fig. 1). To achieve this, we train a neural network to predict cloth edges and corners from a depth image. We also train the network to predict the inner edges, the region interior to the cloth's true edges, for grasp direction estimation. The network is trained on a dataset of RGB-D images extracted from 8 minutes of video of a human manipulating the cloth. The ground-truth for the network is provided by color-labeling the cloth (see Fig. 2), forgoing the need for expensive human annotations. The grasp configuration and uncertainty estimation are both important for grasping the cloth, as mis-estimating the grasp direction and approaching at an angle not orthogonal to the cloth edge mean that grasps are more likely to fail. Using a dense estimate of grasp directional uncertainty, we can choose the grasp point most likely to succeed.

*These authors contributed equally and are listed alphabetically.

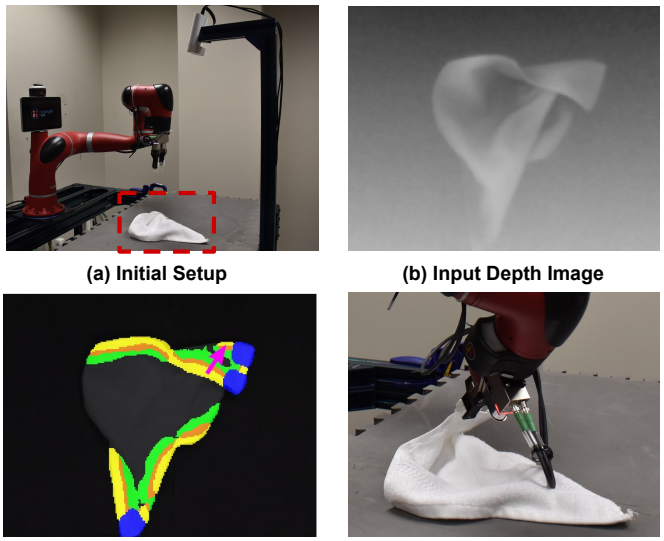


Fig. 1: Grasping using cloth region segmentation: Robot with depth sensor (a) captures depth image of test cloth (b). Depth image is segmented into outer edges (yellow), inner edges (green) and corners (blue) using our cloth region segmentation network (c). Ambiguous regions are colored in orange. Our method selects a grasp location and direction, shown as a magenta arrow. The robot executes a sliding grasp and successfully grips the cloth by its edge.

II. Related Work

A. Cloth Perception

Robotic cloth manipulation is a well-studied domain with a variety of unsolved tasks, including laundry folding [10, 1], laundry unfolding or smoothing [15, 6, 16, 17, 5, 19], bed making [9, 14], and grasping [4, 11, 18].

Many of these approaches use traditional computer vision algorithms to detect cloth regions for various downstream tasks [17, 15, 10]. These perception algorithms usually require significant pre-manipulations to get a more structured configuration of the cloth, thus they are more time consuming than many learning-based methods.

Another category of methods apply learning-based algorithms like YOLO and autoencoder networks for image feature extraction [4, 19]. The most similar method to ours is [14] which learns to identify a corner of a bed sheet by painting a corner red. Our method expands upon this work by predicting a dense segmentation of real edges, inner edges and corners, as opposed to a single 2D corner position. Furthermore, our method outputs dense grasp direction proposals as well as their corresponding uncertainty estimates. We will show that grasp direction proposals and uncertainty estimates are crucial for our grasping performance, enabling us to handle challenging crumpled cloth configurations.

B. Cloth Grasping

Although the focus of our work is on perception rather than grasping, we review prior work on cloth grasping strategies. A simple top-down or angled grasp is commonly used once a grasp point has been selected [17, 14]. A top-down grasp followed by 6DOF grasping on detected corners of the the hanging cloth has also been studied [10].

Other prior works learn a policy for grasping. [11] learns the region in posture parameter space that successful grasps are concentrated. [4] uses Q-learning to train a policy for grasping a folded towel from a stack. [18] uses Soft-Actor-Critic to train a policy for rope and cloth manipulation.

In our work, we identify the area where successful grasps are concentrated in the posture parameter space. Then we execute a hand-designed sliding grasp policy to grasp real edges and corners identified by our perception method.

III. Approach

A. Cloth Region Segmentation

We train a U-Net[13]-based network which receives as input a depth image of the scene containing the cloth. The network predicts semantic labels for each pixel, giving the probability that the pixel contains a cloth outer edge, inner edge, corner, or neither. We then threshold this probability to obtain a semantic segmentation mask for the cloth edge and corner locations (see Fig. 3a for an example of output). To obtain ground-truth labels, we mark all edges and corners with colored paint to get the position of all cloth edges and corners in the image (see Fig. 2). This approach is similar to [14]; differences in our work are explained in Sec. II. We collected 8 minutes of video with a human manipulating this labeled cloth for a total of 6700 RGB-D images split into 6:1:1 train, validation, and test sets.

B. Grasp Configuration Selection

1) Grasp Direction Estimation: To determine the appropriate grasp direction, we augment the above pipeline by also predicting the cloth "inner edges," the 1.5cm region interior to the 1.5cm cloth outer edge. The inner edge labels are shown in green in Fig 2.

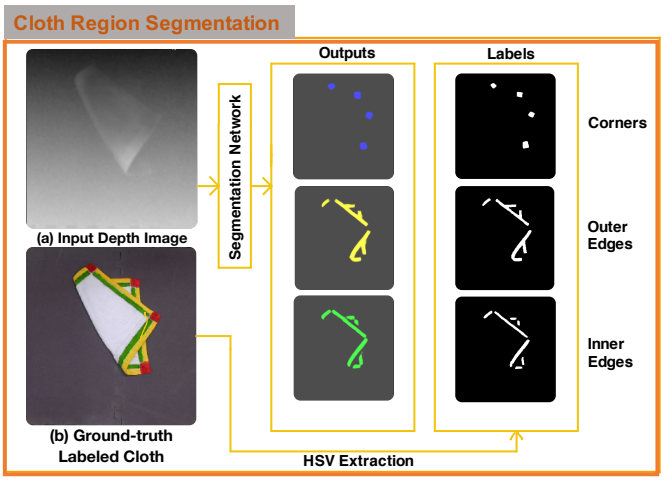


Fig. 2: Training the segmentation network. Input is depth and RGB provides labels.

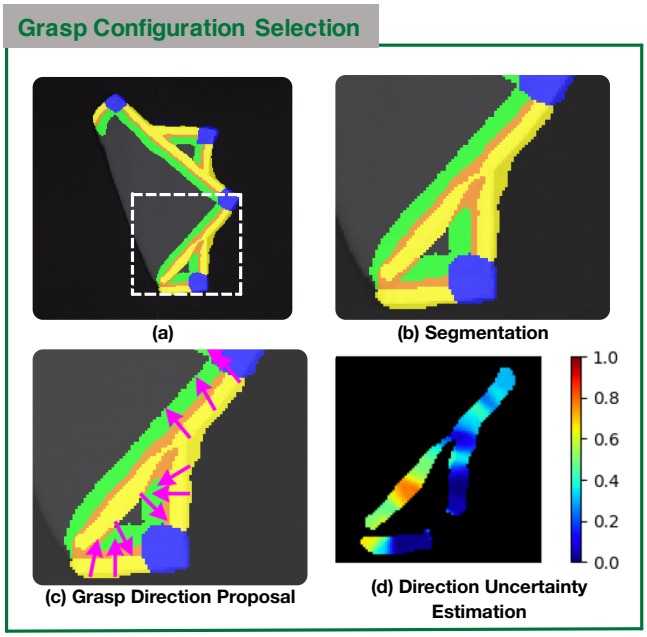


Fig. 3: Illustration of grasp configuration selection. Colors as in Fig. 1. (b) shows the cropped section in (a); (c) shows a subsample of grasp direction proposals for each outer edge points; (d) shows the grasp directional uncertainty for each outer edge points.

Given the predicted segmentation for these cloth regions, we now select a grasp point and direction that allows our sliding grasp policy to most easily grasp the cloth. A sliding grasp that starts with the gripper oriented towards a cloth edge as in Fig. 4 will intercept the edge upon translation. However, a grasp oriented parallel to the edge or approaching from the

reverse direction will not intercept the edge and fail to grasp. Whereas a top-down grasp on overlapping parts of cloth will grasp multiple layers of cloth, a sliding grasp can separate one layer of cloth from another.

We first threshold the output of the network described in Sec. III-A to obtain a set of points estimated to belong to the outer edge $\mathbf{E_O}$ and a set of points that belong to the inner edge $\mathbf{E_I}$. Then, for each outer edge point $\mathbf{p} = [p_x, p_y] \in \mathbf{E_O}$, we find the closest inner edge point $\mathbf{q} = [q_x, q_y]$ by Euclidean distance. With the correspondence between $\mathbf{p}$ and $\mathbf{q}$, we further define the grasp direction at point $\mathbf{p}$ to be the direction along the vector from $\mathbf{p}$ to $\mathbf{q}$. Fig. 3c shows a subset of those grasp directions.

2) Directional Uncertainty Estimation: Fig. 3c shows cases where, due to the complex folds of the cloth, the vector from $\mathbf{p}$ to $\mathbf{q}$ does not indicate an appropriate grasp direction. Thus, for robust grasping, we also compute a measure of the uncertainty in this grasp direction.

We define the uncertainty of the grasp direction for a single point $\mathbf{p}$ to be the variance of the grasp directions predicted by its neighbours. To compute this variance, let $\mathbf{N_k}(\mathbf{p})$ be the set of k closest pixels points in $\mathbf{E_O}$ of $\mathbf{p}$ in Euclidean distance; let α be the angle between $\overrightarrow{\mathbf{pq}}$ and a unit vector along horizontal x axis. Formally we can define the cosine and sine of the grasp direction at $\mathbf{p}$ as

$$f_{cos}(\mathbf{p}) = cos(\alpha) = (q_x - p_x)/\|\mathbf{q} - \mathbf{p}\|_2 \quad (1)$$

$$f_{sin}(\mathbf{p}) = sin(\alpha) = (q_y - p_y)/\|\mathbf{q} - \mathbf{p}\|_2 \quad (2)$$

We can then define observation vectors $\mathbf{x_0}(\mathbf{p})$ and $\mathbf{x_1}(\mathbf{p})$ to contain the cosine and sine of the grasp direction of all points in $\mathbf{N_k}(\mathbf{p})$

$$\mathbf{x_0}(\mathbf{p}) = \left\{ f_{cos}(n) \mid n \in \mathbf{N_k}(\mathbf{p}) \right\} \quad (3)$$

$$\mathbf{x_1}(\mathbf{p}) = \left\{ f_{sin}(n) \mid n \in \mathbf{N_k}(\mathbf{p}) \right\} \quad (4)$$

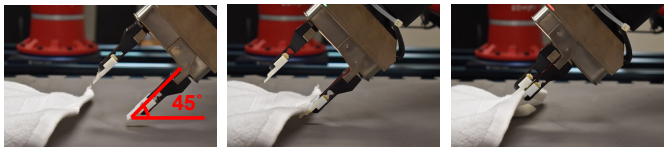
Next we define the sample covariance matrix $\mathbf{K}(\mathbf{p})$ in the usual manner from the observations $\mathbf{x_0}(\mathbf{p})$ and $\mathbf{x_1}(\mathbf{p})$

$$\mathbf{K}_{ij}(p) = \frac{1}{N-1} \sum_{k=1}^{N} \left(x_{ik}(p) - \bar{x}_i(p)\right)\left(x_{jk}(p) - \bar{x}_j(p)\right) \quad (5)$$

where $x_{ij}(p)$ is the jth element of $\mathbf{x_i}(\mathbf{p})$, and $\bar{x}_i(p)$ is the mean of $\mathbf{x_i}(\mathbf{p})$.

Finally, we define the uncertainty of our grasp direction prediction to be the sum of the variances of the individual dimensions, or the trace of $\mathbf{K}$: $Tr(\mathbf{K}(\mathbf{p})) = Var(\mathbf{x_0}(\mathbf{p})) + Var(\mathbf{x_1}(\mathbf{p}))$, where $Var(\mathbf{x_i}(\mathbf{p}))$ is the variance of $\mathbf{x_i}(\mathbf{p})$.

3) Grasp Point Selection: For each outer edge point $\mathbf{p} \in \mathbf{E_O}$, we compute an uncertainty estimate $\mathbf{U}(\mathbf{p}) = Tr(\mathbf{K}(\mathbf{p}))$ as described above. For grasp point selection, we pick the outer edge point $\mathbf{p}$ that has the lowest uncertainty: $\arg\min_{\mathbf{p} \in \mathbf{E_O}} \mathbf{U}(\mathbf{p})$. We use standard projection and motion planning techniques to execute the sliding grasp given the computed grasp configuration.



(a) Pre-slide pose. (b) Post-slide pose. (c) Pinch grasp.

Fig. 4: Sequence of poses for the sliding grasp policy. The sliding action is a translation from the pre-slide to post-slide pose. The slide intercepts the target grasp point on the cloth.

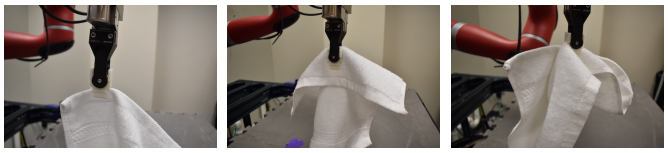
IV. Experiment Setup

All experiments were performed on a 7 DOF Rethink Robotics Sawyer Robot with a Weiss WSG-32 parallel-jaw gripper (see Fig 1a), and a Microsoft Azure Kinect RGB-D sensor. Our test cloth is a white, unlabeled cloth with the same dimensions as the labeled one used for training; however, our depth-based method is color-invariant and generalizes well to different sizes and textures. The video on the website contains examples of such generalization.

V. Experiments

We evaluated our method on grasping cloth edges and corners. Each grasping trial starts with a randomly crumpled cloth in the center of the robot's workspace, obtained by dropping the cloth at least 0.1m from the surface. We then run our method on the robot.

A grasp is considered a success if it pinches a cloth edge or corner and lifts it 30cm above the workspace. Since a grasp could fold over the cloth, we consider a grasp with a single fold over to be a success if the fold is less than or equal to 2cm at its maximum length (see Fig. 5) for edges, and 5cm for corners. All grasps with multiple folds are considered failures.



(a) No fold. (b) Single fold. (c) Multiple folds.

Fig. 5: Examples of cloth grasps. Folds longer than 2cm from edge to fold are considered grasp failures; of these three, only (a) is considered a success.

For the task of grasping cloth edges, we evaluate against three baselines:

- "Segment-Edge" segments the cloth from the table using RANSAC plane fitting. A grasp point is randomly selected from the edge pixels of the segmentation. The grasp direction is determined by the direction of the depth gradient at the selected grasp point.
- "Canny-Depth" applies Canny edge detection [2] to the depth image. The grasp point is sampled uniformly from the set of edge points above an intensity threshold. The grasp direction is determined by the depth gradient direction, as in the above.

- "Canny-Color" is the same as Canny-Depth, except it applies Canny edge detection to the gray-scaled color image. The grasp direction is determined by the color gradient direction instead of depth.

See Fig. B in the appendix for visualizations of these methods.

For the task of grasping cloth corners, we evaluated against the following baselines:

- "Harris-Depth" applies Harris corner detection [7] to the depth image. The maximum intensity value is selected as the grasp point. The depth gradient direction at the grasp point is used to determine the grasping direction, as in the edge grasping experiments.
- "Harris-Color" takes a grayscaled RGB image as input and uses color gradients to determine the grasping direction, but is otherwise the same as the above.

TABLE I: Grasping Cloth Edges and Corners

Method	Edges	Corners
Canny-Depth	0.20 ± 0.00	-
Segment-Edge	0.30 ± 0.00	-
Canny-Color	0.33 ± 0.12	-
Harris-Depth	-	0.05 ± 0.07
Harris-Color	-	0.33 ± 0.15
Our Method	$\mathbf{0.70 \pm 0.20}$	$\mathbf{0.57 \pm 0.06}$

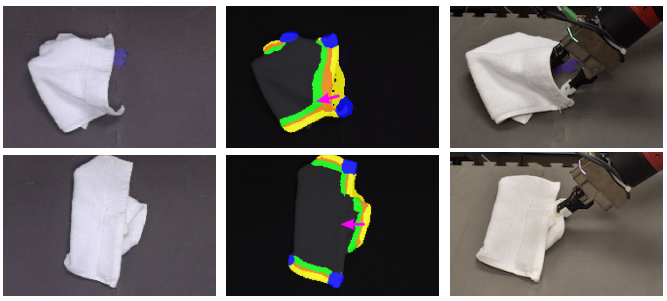
3 trials per method, 10 grasp attempts per trial

The results are shown in Table I. Our method significantly outperforms the baselines in terms of grasp success. The network is largely able to correctly distinguish between edges and folds, determine an appropriate grasp configuration direction, and execute a successful grasp. Averaging over the trials, there were an average of 2.7 failures out of 10 edge grasps due to misdetection, meaning that the grasp point selected was not a real edge. For corners, there were an average of 3 failures out of 10 grasps due to misdetection. There was an average of 0.3 failures out of 10 grasps due to failed grasping. See below for more details on failure cases. For corners, there were an average of 1.3 failures out of 10 grasps due to grasping error. Our method performs worse on corners than on edges. Fewer regions of the image are corners compared to edges, so false positives are more problematic.

However, our method still outperforms the baselines, which perform poorly largely due to an inability to distinguish between real cloth edges vs. folds. Our successful grasps are more often flat with no folding of the cloth, with the edge near horizontal to the gripper tip. Our perception pipeline runs in less than half a second, with no noticeable difference from the baselines.

Failures occurred when the segmentation produced by our method contained errors. Because the cloth is very thin and the depth images captured from our sensor are noisy, the network can fail to get accurate segmentation at cloth edges (see Fig. 6, top row). These segmentation errors affect the grasp selection component that takes the segmentation as input. As a result, we sometimes observed our method selecting grasp points on false positives, which were more likely to result in grasp failures.

Failures also occurred due to grasping areas with valid edges but problematic nearby cloth configurations. For example, overlapping edges can create the appearance of a continuous segmentation, and a grasp on that area will result in grasping both edges (see Fig. 6, bottom row).



(a) RGB Image. (b) Segmentation and Grasp Prediction. (c) Grasp execution.

Fig. 6: Failure cases. (top row) Segmentation bleeds over real cloth edge, leading to poor estimation of grasp height. (bottom row) Grasp fails to avoid grasping nearby folds and edges (note that misdetection has also occurred).

In terms of execution time, the perception component of our method runs in approximately 0.25s, with the segmentation network contributing approximately 0.14s to that total. Grasp execution is a larger bottleneck and requires approximately 15s for all methods.

VI. Conclusion

We present a method to segment real edges and corners of cloth (as opposed to creases or folds) from depth images. Our method also determines a grasp configuration from these segmentations that accounts for directional uncertainty. We demonstrate a system that implements our approach to grasp cloths in crumpled configurations, and we show that our method outperforms various baselines in terms of grasp success rate on grasping success.

Acknowledgments

This work was supported by the National Science Foundation Smart and Autonomous Systems Program (IIS-1849154), the United States Air Force and DARPA under Contract No. FA8750-18-C-0092, LG Electronics, a NSF Graduate Research Fellowship (DGE-1745016), and a NASA Space Technology Research Fellowship (80NSSC17K0233).

References

[1] C. Bersch, B. Pitzer, and S. Kammel. Bimanual robotic cloth manipulation for laundry folding. In *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1413–1419, Sep. 2011.

[2] John Canny. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, (6):679–698, 1986.

[3] David Coleman, Ioan Alexandru Sucan, Sachin Chitta, and Nikolaus Correll. Reducing the barrier to entry of complex robotic software: a moveit! case study. *ArXiv*, abs/1404.3785, 2014.

[4] Satonori Demura, Kazuki Sano, Wataru Nakajima, Kotaro Nagahama, Keisuke Takeshita, and Kimitoshi Yamazaki. Picking up one of the folded and stacked towels by a single arm robot. In *2018 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pages 1551–1556. IEEE, 2018.

[5] Andreas Doumanoglou, Andreas Kargakos, Tae-Kyun Kim, and Sotiris Malassiotis. Autonomous active recognition and unfolding of clothes using random decision forests and probabilistic planning. *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 987–993, 2014.

[6] Kyoko Hamajima and Masayoshi Kakikura. Planning strategy for task of unfolding clothes. *Robotics Auton. Syst.*, 32:145–152, 1997.

[7] Christopher G Harris, Mike Stephens, et al. A combined corner and edge detector. In *Alvey vision conference*, volume 15, pages 10–5244. Citeseer, 1988.

[8] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014.

[9] Michael Laskey, Chris Powers, Ruta Joshi, Arshan Poursohi, and Kenneth Y. Goldberg. Learning robust bed making using deep imitation learning with dart. *ArXiv*, abs/1711.02525, 2017.

[10] Jeremy Maitin-Shepard, Marco Cusumano-Towner, Jinna Lei, and Pieter Abbeel. Cloth grasp point detection based on multiple-view geometric cues with application to robotic towel folding. In *2010 IEEE International Conference on Robotics and Automation*, pages 2308–2315. IEEE, 2010.

[11] Yusuke Moriya, Daisuke Tanaka, Kimitoshi Yamazaki, and Keisuke Takeshita. A method of picking up a folded fabric product by a single-armed robot. *ROBOMECH Journal*, 5:1–12, 2018.

[12] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. dAlché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019.

[13] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[14] Daniel Seita, Nawid Jamali, Michael Laskey, Ajay Kumar Tanwani, Ron Berenstein, Prakash Baskaran, Soshi Iba, John Canny, and Ken Goldberg. Deep Transfer Learning of Pick Points on Fabric for Robot Bed-Making. In *International Symposium on Robotics Research (ISRR)*, 2019.

[15] Dimitra Triantafyllou and Nikos A. Aspragathos. A vision system for the unfolding of highly non-rigid objects on a table by one manipulator. In Sabina Jeschke, Honghai Liu, and Daniel Schilberg, editors, *Intelligent Robotics and Applications*, pages 509–519, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg. ISBN 978-3-642-25486-4.

[16] Dimitra Triantafyllou, Ioannis Mariolis, Andreas Kargakos, Sotiris Malassiotis, and Nikos A. Aspragathos. A geometric approach to robotic unfolding of garments. *Robotics Auton. Syst.*, 75:233–243, 2016.

[17] B. Willimon, S. Birchfield, and I. Walker. Model for unfolding laundry using interactive perception. In *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 4871–4876, Sep. 2011.

[18] Yilin Wu, Wilson Yan, Thanard Kurutach, Lerrel Pinto, and Pieter Abbeel. Learning to manipulate deformable objects without demonstrations. *ArXiv*, abs/1910.13439, 2019.

[19] K. Yamazaki. Gripping positions selection for unfolding a rectangular cloth product. In *2018 IEEE 14th International Conference on Automation Science and Engineering (CASE)*, pages 606–611, Aug 2018.