Regrasping and Unfolding of Garments Using Predictive Thin Shell Modeling

Yinxiao Li[†], Danfei Xu[†], Yonghao Yue[†], Yan Wang[§], Shih-Fu Chang[†§], Eitan Grinspun[†], Peter K. Allen[†]

***Abstract*— Deformable objects such as garments are highly unstructured, making them difficult to recognize and manipulate. In this paper, we propose a novel method to teach a two-arm robot to efficiently track the states of a garment from an unknown state to a known state by iterative regrasping. The problem is formulated as a constrained weighted evaluation metric for evaluating the two desired grasping points during regrasping, which can also be used for a convergence criterion The result is then adopted as an estimation to initialize a regrasping, which is then considered as a new state for evaluation. The process stops when the predicted thin shell conclusively agrees with reconstruction. We show experimental results for regrasping a number of different garments including sweater, knitwear, pants, and leggings, etc.**

I. Introduction

Robotic manipulation of deformable objects is a difficult task especially for sequential manipulations such as unfolding a garment. One particular manipulation task is how to unfold a garment from a random pose to a desired configuration [14][3][4]. This is a difficult task because the states of a deformable object are not easy to track and predict owing to large-dimensional state spaces. In this paper, we propose a system to recognize, manipulate, and unfold a deformable garment into a desired configuration. Key contributions of our paper are:

- A constrained weighted metric for evaluating grasping points during regrasping, which can also be used for a convergence criterion
- A fast, two-stage deformable registration algorithm that integrates simulated results with online localization and uses a novel non-rigid registration method to minimize energy differences between source and target models
- A semantically labeled garment database built with off-line simulation that contains 37 different garments such as sweaters, pants, shorts, dresses, scarves, etc.
- A method for analysis of local curvature using IR scan data for stable grasp
- Experimental results showing that our method applied on a Baxter robot is able to recognize a garment from an arbitrary pose, successfully regrasp it to put it into a known configuration, and place it flat on a table

We are interested in unfolding the garment efficiently as a part of a pipeline for manipulating deformable garments. After several steps of regrasping, the robot holds the garment at two desired positions. Using a sweater as an example, we defined the optimal grasping positions on the two sleeves, respectively. The regrasping is built on the recognition pipeline described in our previous work [10], which applies a volumetric approach to identify the garment pose using a 3D reconstructed mesh model. We extend this previous work by developing a registration-search framework that looks for an optimal position over the entire mesh model. This position is then adopted as a regrasping point in 3-dimensional space and guides the other hand to approach and regrasp. The complete pipeline of a robot folding a garment from a random state is shown in Figure 2, where our work described in this paper spans initial grasping, pose estimation, regrasping for unfolding, and placing the garment flat on a table.

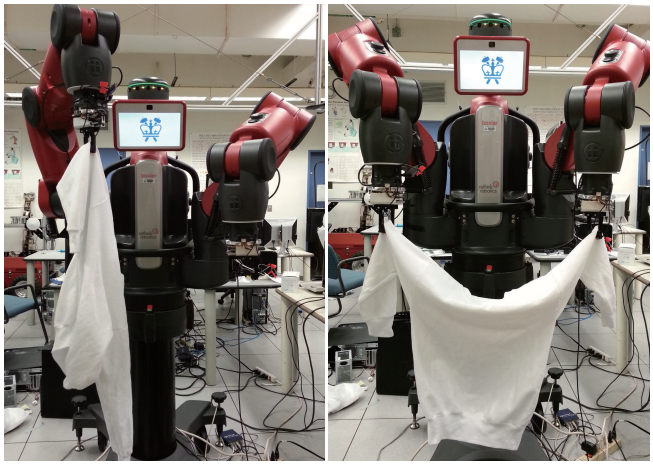


Fig. 1. Our application scenario: a Baxter robot picks up a garment (a sweater here), and a Kinect captures depth images to recognize the pose of the sweater (left). The final unfolding result is shown on the right.

II. Related Work

Kita *et al.* [6][7] have done a series of work on garment pose recognition by grasping at one or more points using a shape matching algorithm. The observation used in this work is just three depth images from top, side, and front views, which may limit its accuracy on more complicated garments. The examples shown in the paper are all relatively rigid garments, lacking further exploration with different materials where deformation is reasonably complex. Willimon *et al.* [18][19] worked on classifying the clothing through interactive perception using a single robotic arm. The heavy dependence on the color-based segmentation, however, makes the method sensitive to the texture variance.

[†] Department Computer Science, Columbia University, New York, NY, USA {yli@cs., dx2143, yonghao@cs., eitan@cs. allen@cs.}columbia.edu

[§] Department of Electrical Engineering, Columbia University, New York, NY, USA {yanwang, sfchang}@ee.columbia.edu

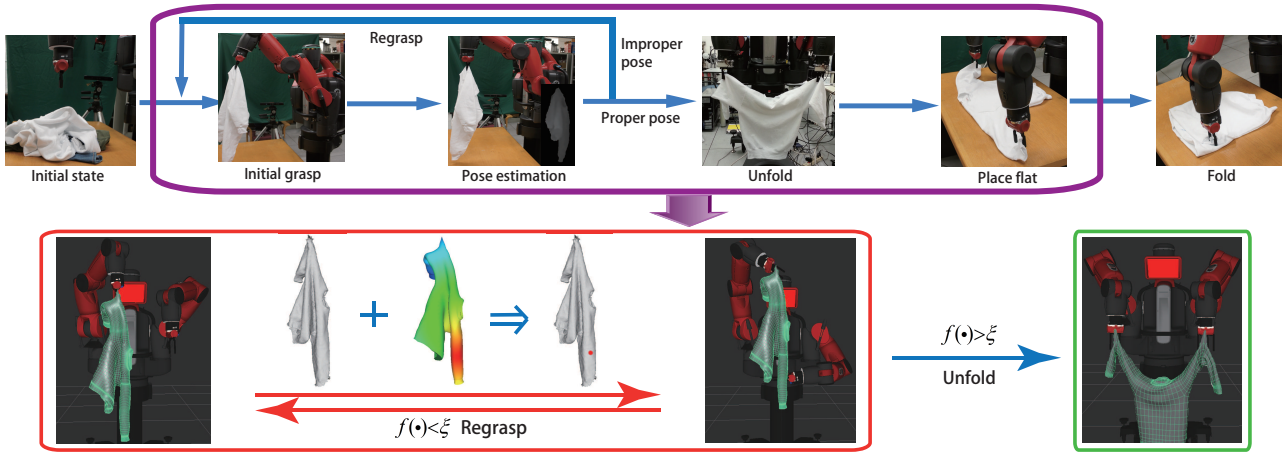


Fig. 2. TOP ROW: The entire pipeline of dexterous manipulation of deformable objects. BOTTOM ROW: If the recognition is not successful, the robot will regrasp the object and repeat the step of pose estimation (the red rectangle).

Wang *et al.* [17], and Miller *et al.* [13] have done work in clothing recognition by manipulation with a PR2 robot. Their method achieves good results but may lack accuracy because of self-occlusion. Our recent work [9] on recognition and pose estimation of deformable clothes uses a Kinect depth sensor. The method recognizes garment category and pose on individual depth images and uses majority voting to get a comprehensive result. A further work [10] first reconstructs a 3D model of the object via a Kinect, and then searches a database of simulated models to predict the pose.

Osawa *et al.* [14] proposed a method using dual-arm to unfold a garment from pick up. However, from the examples in the paper, this method requires a garment with unique color and a clean background for color-based image segmentation. The PR2 robot is probably the first robot that has successfully unfolded deformable objects such as a towel or a T-shirt [12]. The visual recognition in this work targets corner-based features, which may not exist in many soft garments. The subsequent work has improved the prediction of the state of a garment using a HMM framework by regrasping at the lowest corner point [3]. However, this method also requires a clean background, and thus limits the applicability.

The work that is closest to ours is by Doumanoglou *et al.* [4]. This work has impressive results for unfolding of a number of different garments. They use a dual-arm industrial robot to unfold a garment guided by a set of depth images which provide a regrasping point. This method achieves promising accuracy. A major difference between this work and ours is our use of simulated predictive thin shell models for garments to automatically create a large database of garments and their poses. Their training set is a number of physical garments that have been grasped at different grasping points to create feature vectors for learning. Given the physical nature of this training set, it can be very time-consuming to create, and may have problems encompassing a wide range of garments and different fabrics which we can more easily accommodate in the simulation environment. We also use an online registration of a reconstructed volumetric mesh with the simulated model to find regrasping points. By this method, we can choose arbitrary regrasping points without having to train the physical model for the occurrence of the grasping points. This allows us to choose any point on the garment at any time as the regrasping point. We have also developed a closed loop tactile feedback algorithm that allow us to perform a local search at the regrasping point that alleviates the instability of the grasp during regrasping as mentioned in [4].

In our method, the regrasping point is located by mapping the pre-determined point from simulation mesh to the re-constructed mesh. Therefore, a fast and accurate registration algorithm plays a key role in our method. Rigid or non-rigid surface registration is a fundamental tool to find shape correspondence. A thorough review can be found in [15]. Our registration algorithm builds on previous techniques for rigid and non-rigid registrations. First, we use an iterative closest point method [2] to rigidly align the garment. We use a distance field to accelerate the computation. Next, we perform a non-rigid registration to improve the matching by locally deforming the garment. Similar to [8], we find the correspondence by minimizing an energy function that describes the deformation and the fitting.

III. PROBLEM FORMULATION

A. Framework

Our objective is to put the garment into a certain *configuration*[10], such that the garment can be easily placed flat on a table for the folding process. we formulate it as a mathematical optimization problem:

$$\max_{\mathbf{x}_L, \mathbf{x}_R} f(\mathbf{x}_L, \mathbf{x}_R). \tag{1}$$

Here we use two vectors $\mathbf{x}_L, \mathbf{x}_R \in \mathbb{R}^2$ to describe the configuration, i.e. the grasping points on the UV mapping of the garment.[1] and the function f is an evaluation function for such a configuration. We seek a principled way to build

[1]Each garment mesh is defined in a UV 2-dimensional parameter space. The grasping points are chosen from a particular set of UV parameters.

a feedback loop for garment regrasping, which allows us to grasp at pre-determined points on the garment.

For example, the candidate garment is a sweater to be unfold and place flat. A desired configuration is having $(\mathbf{x}_L^*, \mathbf{x}_R^*)$ on the elbows of the sleeves. And our goal is to find a series of regrasping procedures that will converge to a configuration of $(\mathbf{x}_L^*, \mathbf{x}_R^*)$.

B. Optimization Objective

We need a quantitative function defined on the configuration to evaluate its quality. While this can be computed on the continuous surface of the garment, we can also discretize the garment into a set of *anchor points* S_g, which typically contains about 15 points for a garment. After such quantization, the garment pose recognition can be treated as a discrete classification problem, which the current robotics system is able to handle reliably. This also simplifies the definition of the objective function, which then becomes a 2D score table or a matrix, given our robot has two arms.

Although the configuration variables $\mathbf{x}_L, \mathbf{x}_R$ is quantized for the sake of reliable pose estimation, we use a smooth probabilistic Gaussian model to describe the objective function to allow efficient optimization. Specifically, given known locations of grasping points $\mathbf{x}_L, \mathbf{x}_R$, we treat the left and right hand of the robot independently and use the product of the probability of two Gaussians as the objective function:

$$f(\mathbf{x}_L, \mathbf{x}_R) = \mathcal{N}\big(\mathbf{x}_L \mid \mathbf{x}_L^*, \Sigma_L\big) \cdot \mathcal{N}\big(\mathbf{x}_R \mid \mathbf{x}_R^*, \Sigma_R\big), \quad (2)$$

where $\mathbf{x}_L, \mathbf{x}_R \in S_g$. Here $\mathcal{N}(\mathbf{x} \mid \mathbf{x}^*, \Sigma)$ is a 2D Gaussian distribution with $\mathbf{x}^*$ as the mean, and Σ as the covariance on the UV mapping. Typically, $\mathbf{x}_L^*, \mathbf{x}_R^*$ are the target grasping points, and Σs control the tolerance and convergence speed (or number of folding actions to reach the target configuration) of the regrasping process.

In practice, due to the sensor errors of occlusion and deformation, it is not feasible to get a deterministic noise-free observation of $\mathbf{x}_L$ and $\mathbf{x}_R$. Instead, our previous work [10] provides a probabilistic distribution $p(\mathbf{x}_l, \mathbf{x}_r \mid y)$, in which y is the observed mesh model. p is the probabilistic measurement of the confidence given certain grasping points $\mathbf{x}_l, \mathbf{x}_r$. Therefore according to the conditional probabilistic decomposition, the objective can be rewritten as,

$$f(\mathbf{x}_L, \mathbf{x}_R|y) = \sum_{\mathbf{x}_l, \mathbf{x}_r \in S_g} \mathcal{N}\big(\mathbf{x}_l|\mathbf{x}_L^*, \Sigma_L\big)\mathcal{N}\big(\mathbf{x}_r|\mathbf{x}_R^*, \Sigma_R\big)p(\mathbf{x}_l, \mathbf{x}_r|y). \quad (3)$$

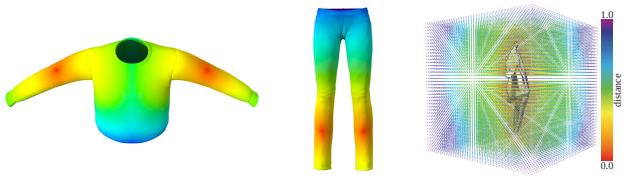
Intuitively, this objective function measures the closeness of each grasping point to the optimal grasping points, scaled by the predicted probability of the current grasping points being accurate. To avoid numerical problems, we do the optimization on $\ln f(\cdot)$, and obtain the objective after substituting the real probabilistic density function of Gaussian:

$$\ln f(\mathbf{x}_L, \mathbf{x}_R) = \sum_{\mathbf{x}_l, \mathbf{x}_r \in S_g} \Big(-(\mathbf{x}_l - \mathbf{x}_L^*)^T \Sigma_L^{-1} (\mathbf{x}_l - \mathbf{x}_L^*) - (\mathbf{x}_r - \mathbf{x}_R^*)^T \Sigma_R^{-1} (\mathbf{x}_r - \mathbf{x}_R^*) + \ln p(\mathbf{x}_l, \mathbf{x}_r|y) \Big). \quad (4)$$

Experimentally, we use the geodesic distance on the garment surface to measure the quality of the regrasping. In our settings, there is no reason to have more weights in one direction against another. Therefore the covariance matrices are set as $\sigma_l I$ and $\sigma_r I$, resulting in the objective function which needs to be maximized as,

$$\ln f(\mathbf{x}_L, \mathbf{x}_R) = -\sum_{\mathbf{x}_l, \mathbf{x}_r \in S_g} \Big(\sigma_l \|\mathbf{x}_l - \mathbf{x}_L^*\|^2 + \sigma_r \|\mathbf{x}_r - \mathbf{x}_R^*\|^2 - \ln p(\mathbf{x}_l, \mathbf{x}_r|y) \Big). \quad (5)$$

The related parameters in the objective such as σ_L and σ_R are set depending on the desired configuration. For example, for sweaters, we set $\mathbf{x}_L^*$ and $\mathbf{x}_R^*$ on the elbow of the two sleeves. The Gaussian formulation ensures a smooth decrease from the expected grasping points, as visualized in Figure 3 as an example.



(a) (b) (c)

Fig. 3. Visualization of the defined objective used in this paper. (a) shows a sweater and (b) show a pair of pants. (c) is the visualization of distance function given a mesh. The color bar on the right shows the normalization distance.

IV. Implemented Method

A. Deformable Registration

After obtaining the location of the current grasp point, we seek to register the reconstructed 3D model to the ground truth garment mesh to establish point correspondences. The input to the registration is a canonical reference ("source") triangle mesh S that has been computed in advance and stored in the garment database, and a target triangle mesh T representing the geometry of the garment grasped by the robot, as acquired by 3D scans of the grasped garment.

The registration proceeds in three steps. First, we scale the source mesh S to match its size to the target mesh T. Next, we apply an iterative closest point (ICP) technique to rigidly transform the source mesh S (i.e., via only translation and rotation). Finally, we apply a non-rigid registration technique to locally deform the source mesh S toward the target T.

Scaling First, we compute a representative size for each of the source and target meshes. For a given mesh, let a_i and $\mathbf{g}_i$ be the area and barycenter of the ith triangle. Then the area-weighted center $\mathbf{c}$ of the mesh is

$$\mathbf{c} = \sum_i^{N_S} a_i \mathbf{g}_i \Big/ \sum_i^{N_S} a_i, \quad (6)$$

where N_S is the number of vertices of the source mesh S. Given the area-weighted center, the representative size l of

the mesh is given by

$$l = \sum_{i}^{N_S} a_i \|\mathbf{g}_i - \mathbf{c}\| \Big/ \sum_{i}^{N_S} a_i. \tag{7}$$

Let the representative sizes of the source and target meshes be l_S and l_T, respectively. Then, we scale the source mesh by a factor of l_T/l_S.

Computing the rigid transformation We use a variant of ICP [2] to compute the rigid transformation. ICP iteratively updates a rigid transformation by (a) finding the closest point $\mathbf{w}_j$ on the target mesh T for each of the vertices $\mathbf{v}_j$ of the source mesh S, (b) computing the optimal rigid motion (rotation and translation) that minimizes the distance between $\mathbf{w}_j$ and $\mathbf{v}_j$, and then (c) updating the vertices $\mathbf{v}_j$ via this rigid motion.

To accelerate the closest point query, we prepare a grid data structure during preprocessing. For each grid point, we compute the closest point on the target mesh using fast sweeping [16], and store for runtime using both the found point and its distance to the grid point as shown in Figure 3, (c). At runtime, we approximate the closest point query for vertex $\mathbf{v}_j$ by searching only among those eight precomputed closest points corresponding to the eight grid points surrounding $\mathbf{v}_j$, thereby reducing the complexity of the closest point query to $O(1)$ per vertex.

After establishing point correspondences, we compute the optimal rotation and translation for registering $\mathbf{v}_j$ with $\mathbf{w}_j$ [2]. We iteratively compute point correspondences and rigid motions until successive iterations converge to a fixed rigid motion, yielding a rigidly registrated source mesh $\bar{S}$.

Non-rigid registration Given a candidate source mesh $\bar{S}$ obtained via rigid registration, our non-rigid registration seeks the vertex positions $\mathbf{v}_j$ of the source mesh S that minimize

$$E_{\bar{S},T}(S) = E_{\text{fit}}(S,T) + E_{\text{def}}(S,\bar{S}), \tag{8}$$

where $E_{\text{fit}}(S,T)$ penalizes discrepancies between the source and target meshes, and $E_{\text{def}}(S,\bar{S})$ seeks to limit and regularize the deformation of the source mesh away from its rigidly registrated counterpart $\bar{S}$. The term

$$E_{\text{fit}} = \sum_{i=1}^{N_S} (\text{dist}(\mathbf{g}_i))^2 \, \bar{A}_i, \tag{9}$$

penalizes deviation of the source and target meshes. Here $\mathbf{g}_i$ is the barycenter of the triangle i, and $\text{dist}(\mathbf{g}_i)$ is the distance from $\mathbf{g}_i$ to the closest point on the target mesh. As in the rigid case, we use the precomputed distance field to query for the distance.

It might appear that the fitting energy E_{fit} could be trivially minimized by moving each vertex of mesh S to lie on mesh T. In practice, however, this does not work because all of the geometry of the precomputed reference mesh $\bar{S}$ is discarded; instead, the geometry of this mesh, which was precomputed using fabric simulation, should serve as a prior. Thus, we introduce a second term to retain as much as possible the geometry of the reference mesh $\bar{S}$:

The deformation term $E_{\text{def}}(S,\bar{S})$, derived from a physically based energy (e.g., see [5]), is a sum of three terms

$$E_{\text{def}}(S,\bar{S}) = \kappa E_{area} + \beta E_{angle} + \alpha E_{hinge}, \tag{10}$$

where α, β and κ are user-specified coefficients.The term

$$E_{area} = \sum_{i=1}^{N_S} \frac{1}{2} \left(\frac{A_i}{\bar{A}_i} - 1 \right)^2 \bar{A}_i, \tag{11}$$

penalizes changes to the area of each mesh triangle. Here A_i is the area of the triangle i, and $\bar{\cdot}$ refers to a corresponding quantity form the undeformed mesh $\bar{S}$. The term

$$E_{angle} = \sum_{i=1}^{N_S} \sum_{k=1}^{3} \frac{1}{6} \left(\frac{\theta_{ik}}{\bar{\theta}_{ik}} - 1 \right)^2 \bar{A}_i, \tag{12}$$

penalizes shearing of each mesh triangle, where θ_{ik} is the kth angle of the triangle i. The term E_{hinge} [5]

$$E_{hinge} = \sum_{e} (\theta_e - \bar{\theta}_e)^2 \|\bar{e}\| / \bar{h}_e, \tag{13}$$

penalizes bending, measured by the angle formed by adjacent triangles. Here θ_e is the *hinge angle* of edge e, i.e., the angle formed by the normals of the two triangles incident to e; $\|\bar{e}\|$ is the length of the edge e, and $\bar{h}_e$ is a third of the sum of the heights of the two triangles incident to the edge e.

We used the secant-version of the L-M method[11] to seek the source mesh S that minimizes the energy Eq.(8). Sample registration results are shown in Figure 4.

B. Grasping Point Localization

We use a pre-determined anchor point (e.g., elbow on the sleeve of a sweater) to indicate a possible regrasping point. The detection of the regrasping point can be summarized in two steps: *global localization* and *local refinement*.

Global localization is achieved by deformable registration. The registered simulation mesh will provide a 3D regrasping point from the recognized state which will be then mapped onto the reconstructed mesh.

In order to improve the regrasping success rate, we propose a step of *local refinement*. The point on the actual garment may be hard to grasp for several reasons. One is that during the garment manipulation steps, such as rotation, the curvature over the garment may change. Another reason is that when considering the width of robot hand gripper, a ridge curve with proper orientation and width should be selected for regrasping. We consider the proper orientation as a direction perpendicular to the opening of the gripper. Therefore, we propose an efficient 1D blob curvature detection algorithm that can find a refined position in the local area over the garment surface via an IR range sensor.

In our experiment, the Baxter robot is equipped with a IR range sensor close to the gripper as shown in Figure 6 top. Once the gripper moves to the same height of the predicted 3D regrasping point from registration, it will perform a horizontal scan search, moving from one side to the other, so that the IR sensor will scan over the full local curvature.

We then apply a curvature detection algorithm that convolves the IR depth signal with a fixed width kernel, where the width is determined by the opening of the gripper. Here we use a Laplacian-Gaussian Kernel $g''(x)$:

$$g''(x) = (\frac{x^2}{\sigma^4} - \frac{1}{\sigma^2})e^{-\frac{x^2}{2\sigma^2}} \quad (14)$$

where x is the depth signal, and σ is the width parameter.

C. Convergence

After the regrasping is finished, we evaluate the current grasping configuration by the objective function $f(\cdot)$. An $f(\cdot)$ greater than a given ξ means the grasping points are on the desired positions, and the robot will then stop regrasping and enter the placing flat mode. The two arms will open to slightly stretch the garment and place it on a table.

V. Experimental Results

To evaluate our results, we tested our method on several different garments such as sweater and pants for multiple trials, as shown in Figure 7 left. A high resolution video of our experimental results is online at `http://www.cs.columbia.edu/~yli/ICRA2015`.

A. Robot Setup

In our experiments, we use a Baxter research robot, which is equipped with two arms with seven degrees of freedom. To improve grasp stability and form a closed loop controller, we add tactile sensors to each of the grippers as shown in Figure 6. Each robot hand is equipped with a IR range sensor which is used for local curvature detection, as described in section IV-B. A Kinect sensor is mounted on a fixed frame at a height of 1.2 meters to capture the depth images.

B. Garment Database and Real-time Recognition

Our recognition method consists of two stages, offline model simulation and online recognition. In the offline model simulation stage, we use a physics engine [1] to simulate the stationary state of the mesh models of different types of garments in different poses. More specifically, we manually label each garment in the database with the key grasping points such as sleeve end, elbow, shoulder, chest, and waist, etc. For each grasping point, we compute the garment layout by hanging under gravity in the simulator. The simulated meshes are used as the training data for 3D shape-based matching and pose recognition [10]. In our experiments, we assume the category of the garment is known. Therefore, we start with the garment pose estimation.

Below we summarize the pose recognition method, details can be found in [10]. We first pick up the garment at a random point. In the online recognition stage, we use a Kinect sensor to capture depth images of different views of the garment while it is being rotated by a robotic arm. The garment is rotated 360° clockwise and then 360° counter-clockwise to obtain about 550 depth images for an accurate reconstruction. We reconstruct a 3D mesh model from the depth image segmentation and volumetric fusion. Then with an efficient 3D feature extraction algorithm, we build up a binary feature vector and finally match against the offline database for pose recognition. One of the outputs is a high-quality reconstructed mesh, which is used for 3D registration and accurate regrasping point prediction, as described below.

Fig. 4. Registration examples. First Row: A sweater grasped at elbow. Second Row:A pair of pants grasped near knee. Each row depicts from left to right: a reconstructed mesh, the predicted mesh from the database, rigid registration only, and rigid plus non-rigid registration.

C. Registration

As described in section III, we apply both rigid and non-rigid registrations. The rigid registration step mainly focuses on mesh rescaling and alignment, whereas the non-rigid registration step refines the results and improves the mapping accuracy. In Figure 4, we compare the difference between using rigid registration only and using rigid plus non-rigid registration side by side. We can clearly see that with non-rigid registration, the two meshes are registered more accurately. In addition, the location of the designated grasping points on the sleeves are also closer to the ground truth points. Note that for the fourth row, after the alignment by the rigid registration algorithm, the state is evaluated as a local minimum. Therefore, there is no improvement by the following non-rigid registration. But as we can see from the visualization, such a case is still good enough for finding point correspondence.

D. Search for Best Grasping Point by Local Curvature

Once we choose a potential grasping point, we can perform a search to find the best local grasping point for the gripper. We are trying to a find a fold in the vicinity of the potential grasping point with a high local curvature tuned to the gripper width that allows for a stable grasp. The opening size of the gripper is approximately $8cm$ and empirically we set $\sigma = 10$ in the equation 14. A plot of its signal, as well as the convoluted signal, are shown in Figure 6 left and right. We can clearly see that the response from the filter is at a minimum where the grasping should take place. The tactile

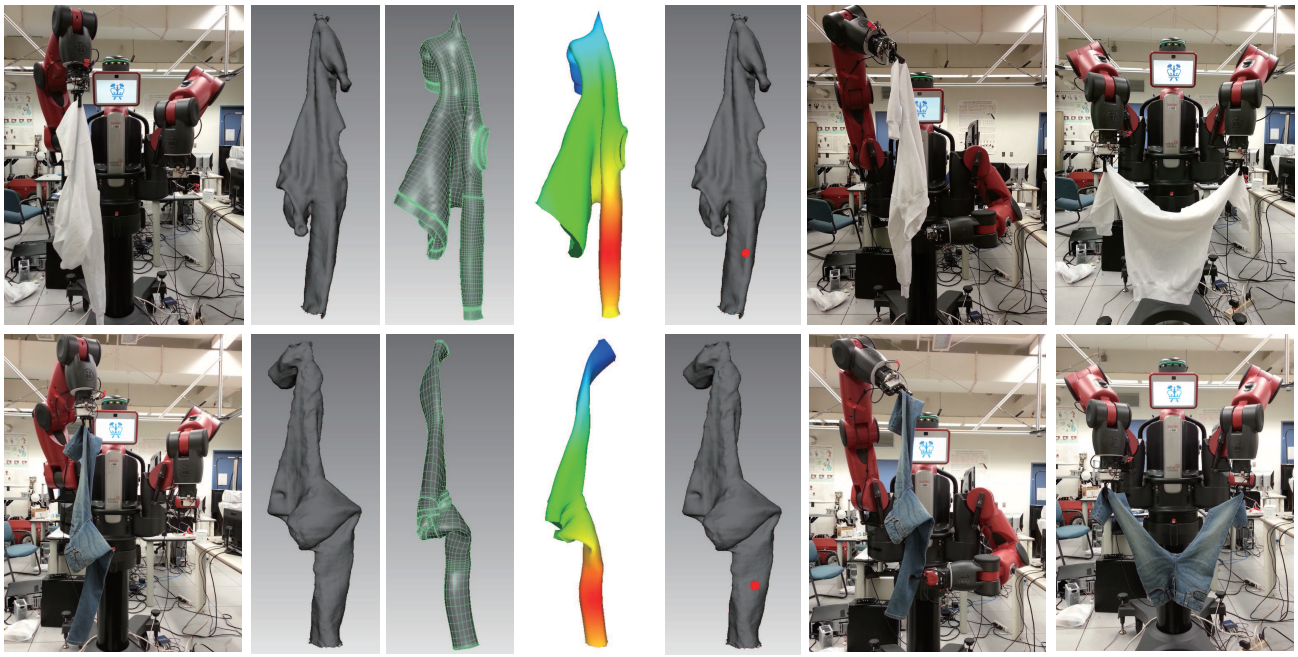


Fig. 5. Examples of each step in our unfolding procedure. For each row from left to right is: a snapshot of initial pick up, a 3D reconstructed mesh, a predicted mesh from database, the predicted mesh with weighted Gaussian distribution distance, predicted regrasping point on the 3D reconstructed mesh, a snapshot of regrasping, and finally a snapshot of unfolding. TOP ROW: The Baxter robot unfolds a sweater following pick up. BOTTOM ROW: The Bater robot unfolds a pair of pants following pick up.

sensors then assure that the gripper has properly closed on the fabric.

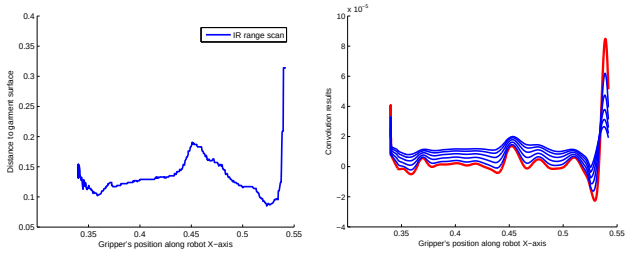


Fig. 6. IR range sensor scan example. LEFT: Single reading plot from the IR range sensor. RIGHT: Convoluted result of the sensor reading and a Laplacian-Gaussian kernel with different kernel size. The lowest point (in red) is the place the gripper should grasp.

E. Iterative regrasping

Figure 5 shows two examples (sweater and pants) of iterative regrasping using the Baxter robot. The robot first picks up a garment at a random grasping point. Once the arm reaches a pre-defined position, the last joint of the arm starts to rotate and the Kinect will capture the depth images as it rotates, and reconstruct the 3D mesh in real-time. After the rotation, a predicted pose is recognized [10] as shown in the third image of each row. For each pose, we have a constrained weighted evaluation metric over the surface to identify the regrasping point as indicated in the fourth image. By registration of the reconstructed mesh and predicted mesh from the database, we can map the desired regrasping point onto the reconstructed mesh. The robot then regrasps by moving the other gripper towards it. With our 1D blob curvature detection method, the gripper can move to the best curvature on the garment and regrasp, which increases the success rate. The iterative regrasping stops when the two grasped points are the designated anchor points on the garment (e.g., elbows on the sleeves of a sweater).

Figure 7 left shows 7 sample garments in our test, and the table on the right shows the results. For each garment, we perform 10 unfolding tests. We have on average an 83% successful recognition rate for the pose of the objects over all the garments. We have on average an 87% successful re-grasping rate for each garments, where regrasping is defined as a successful grasp of the other arm on the garment. 80% of the time we are able to successfully unfold the garment, placing the grippers at the designated grasping points.

Unsuccessful unfolding occurred when either the gripper lost contact with the garment, or the gripper was unable to find a regrasping point. Although we did not perform this experiment, it is possible to restart the method after one of the grippers loses contact as an error recovery procedure.

For the successful unfolding cases, we also report the average number of regrasping attempts. The minimum number of regrasping attempts $= 1$. This happens when the initial grasping is at one of the desired positions, and the regrasping succeeds at the other desired position (i.e., two elbows on the sleeves for a sweater). In most cases, we are able to successfully unfold the garments using $1 - 2$ regraspings.

Among all these garments, jeans, pants, and leggings achieve high success rate because of their unique layout when grasping at the leg position. The shorts are difficult for both recognition and unfolding steps possibly because its ambiguous appearances in different grasping points. One

Garment	# of Trial	Successful Recognition	Successful Regrasping	Successful Unfolding	Avg. # of Regrasps Success Only
Sweatshirt	10	9/10	8/10	8/10	1.6
Sweater	10	8/10	8/10	7/10	1.6
Knitwear	10	9/10	9/10	8/10	1.7
Jeans	10	9/10	9/10	9/10	1.3
Pants	10	8/10	10/10	9/10	1.4
Leggings	10	8/10	9/10	8/10	1.4
Shorts	10	7/10	8/10	7/10	1.9
Average	10	8.3/10	8.7/10	8.0/10	1.6

Fig. 7. LEFT: A picture of our test garments. RIGHT: Results for each unfolding test on the garments. We evaluate the results by recognition, regrasping, unfolding, and regrasping attempts for each test. The last row shows the average of each evaluation component.

observation is that in a few cases, when the recognition is not accurate, our registration algorithm was sometimes able to find a desired regrasping point for unfolding. This is an artifact of the geometry of pant-like garments where the designated regrasping points are at the extreme locations on the garments.

We also show that after grasping at two desired points, the robot will proceed to place the garment on a table. In our experiments, we use cardboard to simulate a table area. As shown in Figure 8, the robot is able to place the garment flat with a simple move when grasping at a pair of two desired grasping points. With such a flat configuration, the robot can begin to fold it, which is part of our future work.

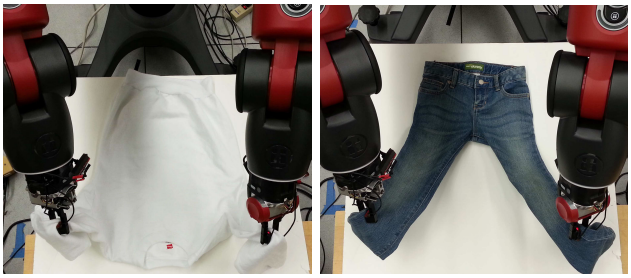


Fig. 8. The Baxter robot places a garment flat on a table. LEFT: The garment is a sweater and the two desired grasping points are on the sleeves. RIGHT: The garment is a pair of pants and the two desired grasping points are on the lower leg parts.

VI. CONCLUSION

In this paper, we propose a novel solution for the problem of unfolding a garment to a desired configuration via regrasping. The unfolding procedure contains garment pick up, pose estimation, regrasping, and placing flat on a table. We use simulated predictive thin shell models for garments to automatically create a large database of garments and their poses, which can be used in a learning algorithm to find object pose. We also create a real-time 3D volumetric model from 3D scanning which can be used for registration with the simulated models. The regrasping point is found by a fast, two-stage deformable object registration algorithm that uses a novel non-rigid registration method to minimize energy differences between source and target mesh models The unfolding state is determined by a constrained weighted objective function for the current two grasping points. Experimental results show that our method is able to unfold a number of common garments with a high success rate.

Acknowledgments We'd like to thank J. Weisz, A. Garg, and J. Augustin for many discussions. We'd also like to thank NVidia Corporation, Intel Corporation, and Takktile LLC for the hardware support. This material is based upon work supported by the National Science Foundation under Grant No. 1217904 and in part by the JSPS Postdoctoral Fellowships for Research Abroad.

REFERENCES

[1] Maya, http://www.autodesk.com/products/autodesk-maya/.
[2] P. J. Besl and N. D. McKay. A method for registration of 3-d shapes. *IEEE Trans. Pattern Anal. Mach. Intell.*, 14(2):239–256, feb 1992.
[3] M. Cusumano-Towner, A. Singh, S. Miller, J. F. OBrien, and P. Abbeel. Bringing clothing into desired configurations with limited perception. In *Proc. ICRA*, 2011.
[4] A. Doumanoglou, A. Kargakos, T-K Kim, and S. Malassiotis. Autonomous active recognition and unfolding of clothes using random decision forests and probabilistic planning. In *Proc. ICRA*, May 2014.
[5] E. Grinspun, A. N. Hirani, M. Desbrun, and P. Schröder. Discrete shells. In *Proceedings of the 2003 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pages 62–67, Aire-la-Ville, Switzerland, Switzerland, 2003.
[6] Y. Kita and N. Kita. A model-driven method of estimating the state of clothes for manipulating it. In *Proc. WACV*, 2002.
[7] Y. Kita, T. Ueshiba, E-S Neo, and N. Kita. Clothes state recognition using 3d observed data. In *Proc. ICRA*, 2011.
[8] H. Li, R. W. Sumner, and M. Pauly. Global correspondence optimization for non-rigid registration of depth scans. In *Proceedings of the Symposium on Geometry Processing*, pages 1421–1430, 2008.
[9] Y. Li, C-F Chen, and P. K. Allen. Recognition of deformable object category and pose. In *Proc. ICRA*, June 2014.
[10] Y. Li, Y. Wang, M. Case, S-F Chang, and P. K. Allen. Real-time pose estimation of deformable objects using a volumetric approach. In *Proc. IROS*, September 2014.
[11] K. Madsen, H. B. Nielsen, and O. Tingleff. Methods for non linear least squares problems. *Tech Report*, 2004.
[12] J. Maitin-Shepard, M. Cusumano-Towner, J. Lei, and P. Abbeel. Cloth grasp point detection based on multiple-view geometric cues with application to robotic towel folding. In *Proc. ICRA*, 2010.
[13] S. Miller, J. Berg, M. Fritz, T. Darrell, K. Goldberg, and P. Abbeel. A geometric approach to robotic laundry folding. *IJRR*, 2012.
[14] F. Osawa, H. Seki, and Y. Kamiya. Unfolding of massive laundry and classification types by dual manipulator. *Journal of Advanced and Intelligent Informatics*, 11(5), 2007.
[15] G. K. L. Tam, Z-Q Cheng, Y-K Lai, F. C. Langbein, Y. Liu, D. Marshall, R. R. Martin, X-F Sun, and P. L. Rosin. Registration of 3d point clouds and meshes: A survey from rigid to nonrigid. *Visualization and Computer Graphics, IEEE Transactions on*, 19(7), July 2013.
[16] Y-H R. Tsai. Rapid and accurate computation of the distance function using grids. *Journal of Computational Physics*, 178(1), 2002.
[17] P-C Wang, S. Miller, M. Fritz, T. Darrell, and P. Abbeel. Perception for the manipulation of socks. *Proc. IROS*, 2011.
[18] B. Willimon, S. Birchfield, and I. Walker. Classification of clothing using interactive perception. In *Proc. ICRA*, 2011.
[19] B. Willimon, I. Walker, and S. Birchfield. A new approach to clothing classification using mid-level layers. In *Proc. ICRA*, 2013.