Bimanual Robotic Cloth Manipulation for Laundry Folding

Christian Bersch, Benjamin Pitzer, Sören Kammel

Abstract— The physical properties of highly deformable objects such as clothing poses a challenging problem for autonomously acting systems. Especially, grasping and manipulation require new approaches that can accommodate for an object's variable and changing appearance.

In this paper, we present a system that is capable of fully autonomously transforming a clothing item from a random crumpled configuration into a folded state. We describe a method to compute valid grasp poses on the cloth which accounts for deformability. Our algorithm includes a novel fold detection and grasp generation strategy, which suggests grasp poses on cloth folds. Machine learning techniques are used to evaluate these grasp poses. In our experiments, we use a stock PR2 robot whose two arms alternately perform grasps on a T-shirt equipped with fiducial markers. The goal of this grasp sequence is to bring the T-shirt into a configuration from which the robot can fold it. In several experiments, we demonstrate the performance of our approach.

I. INTRODUCTION

Since the middle of the 20th century, robots have become a key part of advancing the production industry and are increasingly used in domestic environments today. Laundry is high up on the list of mundane household chores that many people do not want to do and could, presumably, be done by a service robot. In a modern household, doing laundry requires washing, drying, and folding clothing, a three step process. What appears to be easy tasks for a human, poses significant challenges for a robotic system. Most industrial robots were developed to work with rigid objects in a highly repetitive way. Clothes, however, consist of non-rigid materials that cannot be handled with the same repetiveness, which requires a number of additional skills currently unavailable to robots.

The limited manipulation capability of current robots poses a major challenge in automating laundry. Clothes are highly deformable objects. This means that shape, appearance, and other mechanical and visual properties vary as a result of previous handling or as a result of environmental effects such as gravity. In order to manipulate an article of clothing, a robot not only needs to understand the article's current configuration but also needs to estimate how an action would change this configuration.

We consider the task of autonomously folding a T-shirt, where initially the T-shirt is in a random configuration as illustrated in Fig. 1(a). The following subproblems can be identified:

- Picking up the T-shirt from a table.
- Perception of the shirt in its current configuration.

C. Bersch, B. Pitzer and S. Kammel are with Robert Bosch LLC at the Research and Technology Center North America, Palo Alto, CA 94304, USA

- Estimation of where the gripper is holding the shirt.
- Selecting a manipulation action that brings the shirt into a configuration closer to a desired state.
- Executing graps on the T-shirt to physically change its configuration.
- Folding the T-shirt once it has been tranformed into an intermediate configuration from which it can be folded. This configuration should be reachable by repeated execution of the perception, estimation steps and the selected manipulation action.

This paper addresses all of the aforementioned subproblems and we present a system capable of automatically folding T-shirts. Our bimanual manipulation approach alternatingly uses one arm to rotate a piece of clothing in front of the robot's sensors and a second arm to grasp a second point with the goal of bringing the garment into a desired configuration. Grasps are selected by detecting fiducial markers and fold lines on the hanging cloth trying to bring the garment's configuration closer to the goal state. For the particular task of folding a shirt, the goal state of the grasping process is having both grippers holding the shirt at the shoulder points. Grasping poses for a desired grasp point are calculated by simulating the grasping action and evaluating the quality of each grasp pose using a score function that calculates a score based on a set of geometrical features. This score function was automatically trained using machine learning techniques on large set of training data.

II. RELATED WORK

Maitin-Shepard et al. [1] describe a system capable of picking up unseen rectangular towels from a table and detecting and grasping the tips of the towel corners. The towel tips are regrasped until the towel is brought into a configuration where both grippers are holding adjacent towel corners. From this well-defined configuration, the towel can be folded with an open-loop procedure. While their approach works well with rectangular cloth items it may not scale to garments with more complex topologies requiring grasping of arbitrary points on the cloth.

In more recent work, Cusumano-Towner et al. [2], address the subproblem of bringing clothing items into known configurations. They describe an HMM-based method to identify the configuration of a clothing item that is held using two grippers by comparing its observed contour with simulated possible contours. Their focus is on determining the configuration of clothing articles using cues from 2D perception. In contrast to their work, our approach uses 3D and 2D information from multiple views to detect folds in
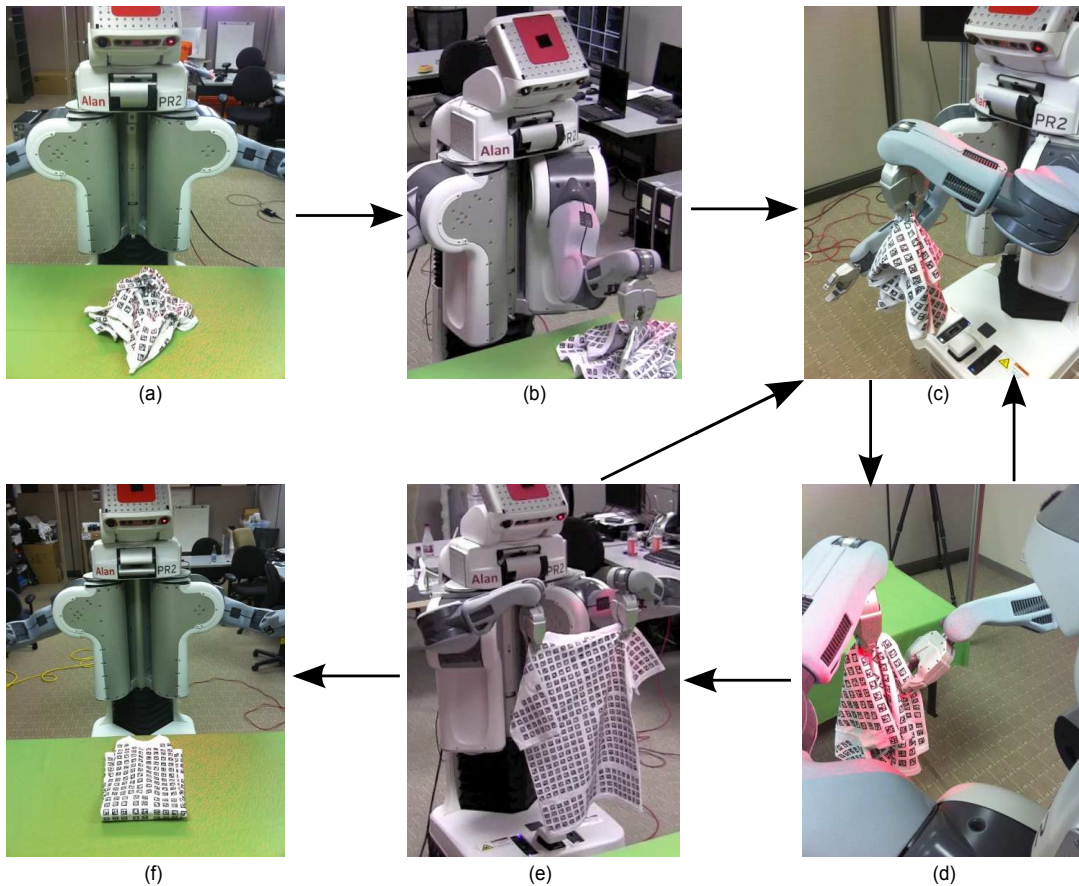
Fig. 1. *T-Shirt folding procedure:* (a) shows the T-shirt in its initial random configuration. After the T-shirt is picked up from the table (b), the robot begins scanning the shirt by rotating it (c). Thereby, the robot acquires point cloud data and records the positions of the fiducial markers. Based on the recorded data, grasps for the free arm are computed, evaluated and the best one according to a greedy policy is executed (d). If the grasp was successful, the other gripper releases the shirt. Steps shown in (c) and (d) are repeated with switching roles of the two arms until both grippers have potentially grasped points that constitute the prespecified desired configuration shown in (e). The correctness of the configuration is verified. If the verification fails, the robot starts over with scanning the shirt (c). If the desired configuration is positively verified, the robot can fold the shirt up. The result is shown in (f).

the cloth configuration such that grasp points are not limited to the cloth's current contour.

Fumaki et al. [3] describe a system with two robotic arms, where, similar to our system, the clothing item is held up by one arm and rotated in order to scan a shirt. They then use the other arm to grasp the lowest hanging point of the item. This process is repeated until the item is in a desired configuration. Note that this requires that both the holding point and the lowest hanging point of the clothing item are inside the reachable space of both arms. This restriction does not apply to our approach as arbitrary points can be grasped. This is essential for handling adult-sized clothing with the PR2 platform used in our experiments, as the distance between two points on the clothing item can exceed the arms vertical or horizontal span.

Furthermore, van den Berg et al. [4] describe a system that folds clothing items that are already spread on a table using gravity-based folds. Their algorithm computes a sequence of motions to fold the item given the geometry of the item and desired fold lines.

The grasp heuristic we describe in this paper is similar to the work of Hsiao et al. [5]. They describe a heuristic to

find grasps for unknown rigid objects based on point cloud data. In contrast their work, an automatically learned score function is used to evaluate these grasps.

Saxena et al. [6] use a Bayesian classifier applied to 2D image features and 3D point cloud features in order to predict whether a grasp on an objects will succeed. As they are dealing with rigid objects, features based on the presence and symmetry of cloud points as well as force closure criteria are used.

Pelossof et al. [7] employ a support vector machine for learning a grasp evaluation function. They simulate grasps of objects with known surfaces and use object shape and grasp parameters as feature vectors to train an SVM that estimates the stability of a grasp. In our approach the surface is not known beforehand and physical grasps are evaluated using geometrical features.

III. SYSTEM OVERVIEW

A. Robot Platform

In our experiments, we use Willow Garage's PR2 robot [8]. The PR2 robot is equipped with two 7-DOF arms which feature two-finger parallel jaw grippers mounted on

a rotating wrist. The PR2 comes with an extensive set of sensors, including wide and narrow angle stereo camera pairs mounted on its head. Furthermore, mono-cameras are located in the robots forearms. The PR2 grippers feature pressure sensor arrays in their fingertips. These were used to determine whether the cloth item has been successfully grasped.

B. Folding Procedure

In the following, we describe the steps that comprise our approach to the problem of bringing a T-shirt from a random configuration into a folded state. The steps involved in this approach are illustrated in Fig. 1.

1) *Pick up:* The robot is confronted with the T-shirt in a random configuration on a table. The T-shirt is detected on a table and its highest point is grasped.

2) *Perception:* The shirt is held up by one arm and rotated by 360 degrees in order to record stereo data and 2D images. During this process, a 3D model of the shirt is generated. This includes a point cloud representation, positions of folds, and positions fiducial markers. This data is then processed in the following steps. See Sec. IV for details.

3) *Estimation of current grasp point:* The current grasp position is estimated based on the position of detected fiducial markers. This process is described in Sec. V.

4) *Selection of next grasp point:* Our current approach uses a greedy policy to select the next grasping action. Given the current grasp point and identifiable regions of the shirt in the current configuration, the goal of this policy is to bring the shirt into a configuration where both shoulders are grasped. The next grasp point selected by the greedy policy is thus the point with the smallest geodesic distance to target grasp points (i.e., the shirt shoulders). The geodesic distances are calculated from a mesh representation of the T-shirt as shown in Fig. 2(b).

5) *Grasp pose computation and evaluation:* For the selected grasp point, a pose for the gripper is computed and evaluated. See Sec. VI for details. If for a selected grasp point no valid pose can be found, the next best point according to step 4) is picked. This process repeats until the best valid and executable grasp pose is found.

6) *Grasp execution:* The determined grasp pose is executed. The T-shirt may be rotated by the holding gripper such that the other grasping gripper can reach the computed pose.

7) *Grasp verification:* If the desired goal configuration is believed to be reached a verification step is performed. The grasp points of both grippers are estimated and the grippers are rotated to ensure the T-shirt is hanging correctly between both grippers. If the shirt is not grasped at its shoulders, the procedure continues at step 2).

8) *Folding:* Once the T-shirt is held at the desired target points located at the shoulders, an open loop folding
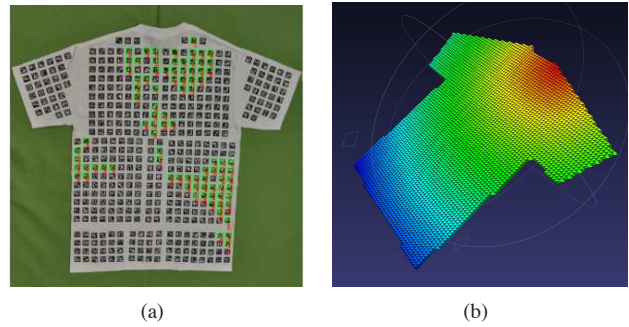

(a) (b)

Fig. 2. (a) The T-shirt with printed fiducial markers. The markers that were detected during one 360-degree scan are highlighted in the image. (b) A triangle mesh is generated for measuring geodesic distances between positions on the T-shirt surface. Here, the geodesic distance of points on the shirt surface to the left shoulder is visualized.

routine is initiated. Thereby the sleeves and shirt sides are rotated inwards and the shirt is placed on a table with one folding motion. This simple open loop routine yields satisfactory results.

The perception step 2), the grasp point estimation step 3) and the grasp computation and evaluation step 5) form the key components of our system and are explained in detail in the next three sections.

IV. PERCEPTION

A. Fiducial Markers

In our current system, we focus on the grasping and planning aspects of cloth manipulation. Therefore, we chose to sidestep the problem of identifying the part of a cloth item to which a particular fold or section belongs. However, interesting work to solve this problem has been presented in [2]. In order to identify the cloth parts, fiducial markers which contain digital codes as ID numbers as described in [9] were printed onto a regular adult-sized T-shirt. We modified a version of ARToolKitPlus [10] to detect multiple markers on the cloth simultaneously under varying lighting conditions. The marker layout is shown in Fig. 2(a). From pre-recorded photographs of the front and rear side of the flat shirt, a triangle mesh representation is generated automatically. Only the openings in the sleeves, collar and bottom are manually annotated.

This mesh representation, as shown in Fig. 2(b), allows for fast calculation of geodesic paths between points on the T-shirt surface using the algorithm described in [11]. The length of these geodesic paths are used for grasp point estimation.

B. Point Cloud Acquisition and Preprocessing

In order to find and evaluate grasps on the clothing item, a precise representation of its 3D configuration is needed. Thus, the robot suspends the item from its vertically aligned gripper such that it can rotate the hanging cloth without inducing significant changes to the cloth configuration. This effectively allows for multiple camera view points the current cloth configuration. While rotating, point clouds from the stereo camera are recorded and registered as shown in

Fig. 3(b). During this scanning process, the mono cameras in the robot's head and arms detect the 3D positions of fiducial markers. The free arm is moved up and down to increase the number of detected markers. Additionally, camera images from a number of distinct positions are taken for detecting folds based on 2D features.

V. ESTIMATION OF GRASPED POINTS

During the scanning process, the 3D positions $\boldsymbol{x}_i$ of the detected markers i as well as the pose of current gripper $\boldsymbol{x}_g$ are recorded. This information is used to calculate a maximum-likelihood estimate of the T-shirt mesh vertex $\boldsymbol{v}_g^*$ at which the shirt is currently held by the gripper. Similarly, this estimate is used to verify that both shoulders have been grasped, by detecting markers when the shirt is held up.

Similar to [4], we assume infinite flexibility of cloth material, i.e., there are no internal forces in the cloth as a result from bending it. Also, we neglect the effect of stretching of the cloth due to its own weight and friction between cloth parts. Thus, the euclidean distance between each pair of points on the cloth can be at most the length of the geodesic path between these points.

When the cloth item is suspended from the gripper without contact to other objects or the robot itself, we can make the observation that all points of the clothing piece tend to fall straight down due to gravity (or at least on a vertical cone centered at the gripper position due to the thickness of the material). Thus, the observed euclidean distance between the position of holding gripper's fingertips and a given marker position $d_{i_{\text{eukl}}} = ||\boldsymbol{x}_i - \boldsymbol{x}_g||_2$ should be equal to the geodesic distance $d_{i_{\text{geod}}} = geodesicDistance(\boldsymbol{v}_i, \boldsymbol{v}_g)$ between this marker's surface position on the cloth and a surface position v_g where the T-shirt may have been grasped.

Furthermore, we make the naive-Bayes assumption that the detections of the markers are conditionally independent. Thus, the ML-estimate of $\boldsymbol{v}_g$ can be expressed as

$$\boldsymbol{v}_g^* = \operatorname*{argmax}_{\boldsymbol{v}_g} \prod_i p(d_{i_{\text{eukl}}} | d_{i_{\text{geod}}}) \quad,$$

where we choose $p(\cdot)$ to be a Gaussian density function with mean $d_{i_{\text{geod}}}$ and a standard deviation $\sigma(d_{i_{\text{geod}}})$ which increases linear with the $d_{i_{\text{geod}}}$. Equivalently, we can minimize the resulting negative log-likelihood:

$$\boldsymbol{v}_g^* = \operatorname*{argmin}_{\boldsymbol{v}_g} \sum_i \frac{(d_{i_{\text{eukl}}} - d_{i_{\text{geod}}})^2}{\sigma(d_{i_{\text{geod}}})^2} \quad.$$

We increase σ to address violations of the above mentioned assumptions as they become more pronounced for larger geodesic distances. In particular, the infinite flexibility assumption does not always hold, as topologically distant cloth parts may be bend upwards and remain located on top of other cloth parts, causing observed euclidean distances to be much smaller than predicted. As these cases only appear if there is enough cloth material to bend, the geodesic distance is typically large and thus the effect on the grasp point estimate is in practice negligible.

VI. GRASP POSE COMPUTATION AND EVALUATION

Finding and evaluating grasps for highly deformable objects such as suspended cloth differs significantly from grasp planning for rigid objects. While for rigid objects the torques and forces exerted onto the object are of major concern [12], grasp planning for cloth items must take the possible flexing of the material into account.

For instance, grasps on the cloth may be desired that require the sections of the cloth to be pushed away during the grasp approach. This type of grasp would be impossible with traditional grasp planning approaches for rigid object because the robot's end effector is in collision state with the portion of the cloth that can be pushed away. Thus, the grasp may still be valid, even though the grasp is in collision with the object in the current static configuration.

Here, our goal is to find whether it is possible to grasp a clothing item close to a given position on the T-shirt manifold, and if so, return a valid 3D grasp pose that corresponds to this surface position. In the following, we present a heuristic approach to solve this problem using both 3D stereo data and 2D image features.

A. Fold Detection

For planning grasp poses on garnments, it is important to distinguish between the part of the cloth item which is to be grasped and the parts that can be ignored or moved away. For this, fold lines that separate these parts are identified based on image gradients caused by the folds' shadows. These fold lines are then used to segment the point cloud as illustrated in Fig. 3(b). Furthermore, the segmentation allows for further processing of a largely reduced point cloud, which results in significant speed gains for the grasp computation. In this paper, we focus on detecting folds that run in a ray within a certain offset from the gripper, as these folds are the most commonly encountered.

Two neighboring folds or overlapping portions of the cloth create a distinct shadow line on the shirt surface. These lines can be reliably detected in the camera images by filtering the shadow gradients and applying a Hough transform as described in [13]. Information from the detected fiducial markers was not needed, even though this information would allow to filter out false positives, which were mostly caused by the marker pattern itself. To remove false positives, we check for the spacing of gaps in the detected lines that occur at roughly the spacing of the markers. A resulting detection is shown in Fig. 3(a).

The detected 2D lines in the camera image are then projected onto the cloth surface. For this, we use RANSAC [14] to approximate tangent planes to cloth surface. The 3D fold lines are obtained by projecting the 2D lines onto their corresponding tangent planes.

Given a desired point to grasp $\mathbf{p}_g$, a point cloud $P = \{\mathbf{p}_i\}$ and a set of 3D lines $L = \{l_j\}$, where $l_j = \mathbf{a}_j + t \cdot \mathbf{b}_j$, we determine the cloud points belonging to the cloth portion that we want to grasp as follows:

First, the two lines l_k and $l_m \in L$ closest to $\mathbf{p}_g$ are determined, subject the constraint that the perpendiculars
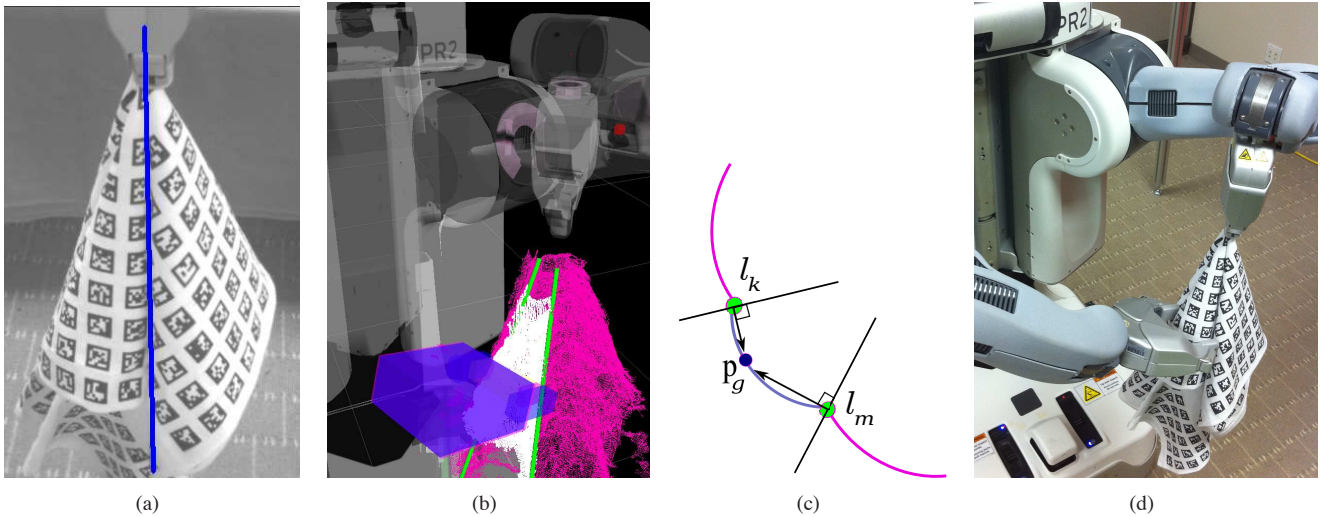
Fig. 3. (a) shows a detected fold line on a 2D image. In (b) the resulting 3D lines are shown in green. The point cloud segment relevant for grasping is colored in white, other points are shown in red. The blue box-based gripper model is used for grasp simulation. The drawing in (c) illustrates conceptually from a top view perspective how the tangent planes were constructed which segment the point cloud. In (d) the resulting physical grasp is shown.

$\mathbf{r}_k(\mathbf{p}_g)$ and $\mathbf{r}_m(\mathbf{p}_g)$ from $\mathbf{p}_g$ to l_k and l_m have opposing directions. The perpendicular $\mathbf{r}_j(\mathbf{p}_i)$ from a point $\mathbf{p}_i$ to a line $l_j = \mathbf{a}_j + t \cdot \mathbf{b}_j$ is

$$\mathbf{r}_j(\mathbf{p}_i) = \mathbf{a}_i - \mathbf{p}_i + \frac{(\mathbf{p}_i - \mathbf{a}_i)^T \mathbf{b}_i}{\mathbf{b}_i^T \mathbf{b}_i} \cdot \mathbf{b}_i \quad .$$

Thus, we can find the first line l_k for which the perpendicular distance $||\mathbf{r}_k(\mathbf{p}_g)||_2$ is minimal. Equivalently, the second line l_m can be chosen from the set of lines $\{l_r \in L : \mathbf{r}_k(\mathbf{p}_g)^T \mathbf{r}_r(\mathbf{p}_g) < 0\}$ for which its perpendicular is minimal if this set is not empty.

These perpendiculars are used to decide for each point in P whether it belongs to the cloth portion we want to grasp. We compare the perpendicular from the desired grasp point $\mathbf{p}_g$ to the perpendicular from each point $\mathbf{p}_i$ in the point cloud. Again, we are interested in their relative orientation, so we calculate dot product $\mathbf{r}_k(\mathbf{p}_i)^T \mathbf{r}_k(\mathbf{p}_g)$. If it is negative, $\mathbf{p}_i$ must lie on the other side of the l_k and is thus not considered part of the cloth portion where $\mathbf{p}_g$ is located on. Effectively, this cuts the point cloud along the plane passing through the 3D line l_k with a normal vector located on this line pointing towards $\mathbf{p}_g$ as illustrated in Fig. 3(c). The same procedure is applied using the second line l_m. Note, that theoretically there are "degenerate" cases conceivable where this segmentation rule does not discriminate accurately between points on the cloth part on which $\mathbf{p}_g$ is located and the neighboring part separated by a fold, e.g., if the neighboring segment wraps around the cloth portion to be grasped and $\mathbf{p}_g$ is close to the fold. However, we did not observe this effect in practice, and it is negligible compared to missed fold detections and missing surface points. Also, we are only interested in segmenting the cloth parts around the desired grasp point $\mathbf{p}_g$ as for the grasp pose computation we limit the radius of search to feasible poses, i.e. how the planes segment the point cloud on the far side of the cloth surface.

B. Grasp Generation

In the previous section, we described how we obtain a point cloud segment surrounding a desired grasp position. We now use this to generate a set of valid grasp poses for a two-finger gripper. We thereby extend the approach described in [5], where the problem of computing and ranking grasp poses for unknown objects by fitting a simplified gripper model around the point cluster belonging to an object is addressed. However, unlike this work, our algorithm allows for arbitrary grasp approach directions and uses machine learning techniques to evaluate grasps as described in Sec. VI-C

As segments of folds which are not in the current area of interest for grasping are virtually cut away from the point cloud, collision checking on the remaining point cloud can be used to generate valid grasps. In our algorithm, a gripper model consisting of three collision boxes is used: One for each gripper finger and one for the palm of the gripper. When fitting the gripper model to the cloth part, none of its cloud points may lie within one of the collision boxes in order to constitute a valid grasp pose. The space between the finger boxes is represented by forth box which must contain cloud points.

Given a position to grasp, we randomly sample positions and orientations around the given position using a rapidly space exploring random tree [15]. We initialize the search with the closest cloud point and its corresponding normal. The surface normals is estimated as described in [16] by fitting a least-square plane through a set of neighboring points within a certain radius, where the plane's normal is used to approximate to the point's surface normal. For markers close to the natural border of the shirt, we additionally initialize a random tree using a vector orthogonal to the surface normal. The sampled gripper pose is locally adjusted by moving the simulated gripper pose along its roll axis, such that it is not in collision, but the number of points in the space between the fingers is maximized. Thereby, the gripper pose is also

adjusted in lateral direction and as well as its opening width, so that the gripper is centered between the left and rightmost points inside the space box and the fingers touch them.

C. Grasp Evaluation

As a set of valid grasp poses for a given 3D point is calculated, the individual grasp must be analyzed and the likelihood of their success evaluated. A score function that assigns each grasp a score based on a set of geometrical features was generated using a support vector machine with a Gaussian kernel function [17].

The features are based on the position and normal vectors of the cloud points inside the space box of the gripper model. The following features were selected:

- The total number of points in the space box.
- The number of space box points within a margin of each finger: Typically, the more points are close to the fingers contact surface the more robust the grasp.
- The orientation of the normals of the points within a margin of each finger: Ideally, the object's surface normals are pointing towards the finger contact surface which will be touching them during a grasp. As a numerical features, the average of the dot products of the surface normals within prespecified regions of the spacebox and the normals of the closer finger's contact surface is used.
- The distance between the fingertips and the closest point to the palm in longitudinal gripper direction.
- The grippers opening width: As the number of points in the space box is not independent from the grippers opening width this feature was included.

During the training phase, the robot randomly picks points on the cloth surface and physically executed the first valid grasp found. For each such grasp, the feature vector is recorded and a label indicating whether or not the grasp would have been successful. The success of a grasp is verified using the fingertip pressure sensor array and the encoder reading of gripper opening width. The whole process for recording the training set can thus run fully automatically.

VII. EXPERIMENTS

A. Grasp Evaluation Learning

In order to generate training data for estimating the grasp evaluation function, we let the robot continuously grasp random points on the T-shirt while being suspended from one gripper. Successes and failures were automatically recorded by analyzing the fingertip pressure sensors and gripper opening positions. If a grasped succeeded, the T-shirt was suspended from the grasping gripper and the other gripper released the shirt and was subsequently used for grasping attempts. In total, 487 sequential grasp attempts were recorded, 271 of which being successful and 216 failed. We trained a SVM on this data set using 10-fold cross validation. For successful grasps we achieved an F-measure of 77% as the harmonic mean of a precision of 76% and a recall of 78% which we chose as operating point. The plot of true positive rates vs. false positive rates are shown in the
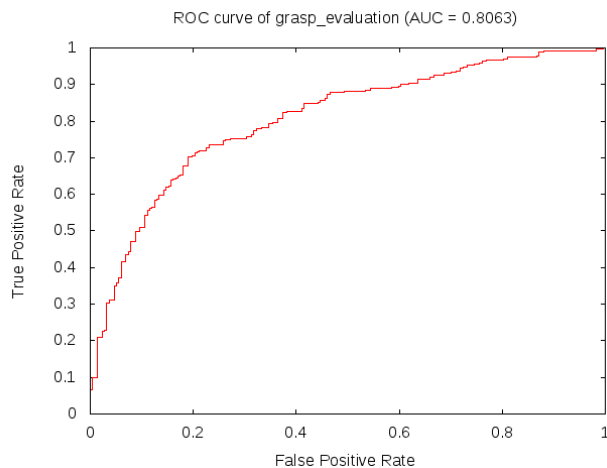

Fig. 4. True positive vs. false positive rate for grasp classification resulting from 487 training examples using 10-fold cross validation.

ROC-curve in Fig. 4. Note that for the experiments for the complete folding procedure the actual success rate of grasps in terms of precision was higher than predicted[1]. This effect may be explained by the fact that for each grasp point a set of grasps is calculated and only the one with the highest score, i.e. the highest decision value is executed.

B. Folding Procedure

For a complete folding procedure starting with picking up the shirt from the table to folding of the shirt, 20 complete runs were conducted. The results are shown in Tab. I.

TABLE I

FOLDING PROCEDURE RESULTS

	Average	Min	Max
Total number of grasps per run	5.9	1	15
Percentage of failed grasps	0.14	0.00	0.20
Total duration per run	19mins 6s	5min 45s	45min 58s

18 runs completed successfully. However, we observed two failures. In one case the shirt got mechanically caught behind the gripper fingers when it tried to release it and human intervention was required. This was the only time this failure case was observed including the 487 grasps during the evaluation function training procedure. The second failure case occurred when the grasp verification predicted a successful grasp even though one gripper had caught multiple layers of cloths. Thus, the shirt was not hanging correctly from both grippers and the folding routine failed.

We observed widely varying counts of regrasps until the shirt was in the desired configuration. In one case the shirt was initially picked up from the table at the shoulder such that only one grasp of the other arm was necessary to bring the shirt into the desired configuration. At most we observed 15 regrasps. Part of this can be explained by the

[1]We can only observe precision. False or true negatives do not occur because a grasp predicted to fail at given operating point will not be executed.

sub-optimality of the greedy heuristic. This policy often yields reasonable results because it brings the shirt quickly in a configuration where grasp points close to shirt shoulders can be reached even if the shirt is initially hanging from its bottom. However, grasping of points very close but not close enough to the target positions turned out be detrimental to the performance as it prevented the actual target positions to be grasped. Grasping of points further away may lead to a reduced number of regrasps.

The average duration to complete the full process was 19mins 6s. Most of the time is spent on the scanning process which takes about 2mins 20s for each grasp. This is mostly due to the need of moving the free arm to capture images from different arm camera poses. Additional fixed cameras mounted on the robots torso would have a similar effect while reducing the scan time tremendously. Also, scanning only interesting parts of the cloth configuration would be conceivable but this would require a more elaborate planning approach.

VIII. Summary and Conclusion

In this paper, we presented a robotic system capable of autonomously folding a T-shirt that is placed onto a table in front of the robot in an arbitrary configuration. The folding procedure first brings the shirt into a defined configuration, e.g. grasping it with two grippers on the shoulder parts next to the collar, and subsequently folds it into a stack. Since the folding itself can be performed rather easily with an open-loop algorithm, we put the emphasis of our work on bringing the shirt into a suited configuration. For this, it is necessary to determine the shirt's *current* configuration and plan a sequence of grasps in order to bring it into a well-suited state. Currently, a greedy strategy for picking the next grasping point is used. In order to perform a grasping operation at the desired shirt position, grasping actions were simulated and the quality of each grasp pose was evaluated using a function that calculates a score based on a set of geometric features. This score function was automatically trained using a SVM. Finding good poses for grasping is especially difficult for deformable objects such as cloth. Most planning algorithms treat the object to grasp as rigid and would reject poses that move the gripper into a fold. For this purpose, a novel fold detection and grasping strategy was developed that in a preprocessing step virtually cuts clear the fold to grasp. In all our experiments we used a stock PR2 robot whose two arms alternatingly perform grasps on a T-shirt equipped with fiducial markers.

For the whole folding procedure, starting from picking up a T-shirt in a random configuration to folding and placing it on a table we observed a success rate of 90%. To further improve these results, we plan to replace the greedy strategy that is used to determine the next grasping point by a learned policy that is capable of choosing a grasping point geometrically further away from the goal position to overcome regrasping problems as described in the last chapter.

The described procedure is likely to apply to general garments, as, except for the open-loop folding routine, no T-shirt specific knowledge is incorporated. The use of fiducial markers allowed us to focus on grasping and manipulation strategies. However, in future work we intend to replace the markers with 2D and 3D vision-based approaches for detecting and identifying keypoints on off-the-shelf garments.

References

[1] 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 *ICRA*, 2010, pp. 2308–2315.

[2] M. Cusumano-Towner, A. Singh, S. Miller, J. F. O'Brien, and P. Abbeel, "Bringing clothing into desired configurations with limited perception," in *Proceedings of IEEE International Conference on Robotics and Automation (ICRA) 2011*, May 2011.

[3] F. Osawa, H. Seki, and Y. Kamiya, "Unfolding of massive laundry and classification types by dual manipulator," *JACIII*, vol. 11, no. 5, pp. 457–463, 2007.

[4] J. van den Berg, S. Miller, K. Goldberg, and P. Abbeel, "Gravity-based robotic cloth folding," in *9th International Workshop on the Algorithmic Foundations of Robotics (WAFR)*, 2010.

[5] K. Hsiao, S. Chitta, M. Ciocarlie, and E. G. Jones, "Contact-reactive grasping of objects with partial shape information," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 10/2010 2010, pp. 1228–1235.

[6] A. Saxena, L. L. S. Wong, and A. Y. Ng, "Learning grasp strategies with partial shape information," in *in AAAI*, 2008.

[7] R. Pelossof, A. Miller, P. Allen, and T. Jebara, "An svm learning approach to robotic grasping," in *In IEEE International Conference on Robotics and Automation*, 2004, pp. 3215–3218.

[8] K. Wyrobek, E. Berger, M. Van der Loos, and K. Salisbury, "Towards a personal robotics development platform: Rationale and design of an intrinsically safe personal robot," in *ICRA*, 2008, pp. 2165–2170.

[9] M. Fiala, "Artag, a fiducial marker system using digital techniques," in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 2 - Volume 02*, ser. CVPR '05. Washington, DC, USA: IEEE Computer Society, 2005, pp. 590–596.

[10] "ARToolKitPlus." [Online]. Available: https://launchpad.net/artoolkitplus

[11] R. Kimmel and J. A. Sethian, "Computing geodesic paths on manifolds," in *Proc. Natl. Acad. Sci. USA*, 1998, pp. 8431–8435.

[12] A. T. Miller and P. K. Allen, "Graspit! a versatile simulator for robotic grasping," *Robotics & Automation Magazine, IEEE*, vol. 11, no. 4, pp. 110–122, 2004.

[13] J. Matas, C. Galambos, and J. Kittler, "Robust detection of lines using the progressive probabilistic hough transform," *Computer Vision and Image Understanding*, vol. 78, no. 1, pp. 119 – 137, 2000.

[14] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[15] S. M. Lavalle, "Rapidly-exploring random trees: A new tool for path planning," Tech. Rep., 1998.

[16] J. Berkmann and T. Caelli, "Computation of surface geometry and segmentation using covariance techniques," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, pp. 1114–1116, November 1994.

[17] C.-C. Chang and C.-J. Lin, *LIBSVM: a library for support vector machines*, 2001, software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.