Towards Robotic Garment Folding: A Vision Approach for Fold Detection

David Estevez, Juan G. Victores, Santiago Morante and Carlos Balaguer

***Abstract*— Folding clothes is a current trend in robotics. Previously to folding clothes, they have to be unfolded. It is not realistic to perform model-based unfolding, as every garment has a different shape, size, color, texture, etc. In this paper we present a garment-agnostic algorithm to unfold clothes that works using 3D sensor information. The depth information provided by the sensor is converted into a grayscale image. This image is segmented using watershed algorithm. This algorithm provide us with labeled regions, each having a different height. In this labeled image, we assume that the highest height region belongs to the fold. Starting on this region, and ending in the garment border, tentative paths are created in several directions to analyze the height profile. For each profile, a bumpiness value is computed, and the lowest one is selected as the unfolding direction. A final extension on this line is performed to create a pick point on the fold border, and a place point outside the garment. The proposed algorithm is tested with a small set of clothes in different positions.**

I. INTRODUCTION

Folding clothes is a common and necessary, but tedious, task for humans. Additionally, due to the increasing aging of the world population, a growing need exists for robots to be able to help us with laundry. However, working with non-rigid objects such as clothes is a difficult task for robots, due to the complexity of modeling and manipulating deformable, thin objects. Clothes can be easily entangled when doing laundry, and recognizing individual garments and their category just from color or depth image analysis becomes an almost impossible task, due to occlusions amongst the cluttered clothes. Another challenging aspect when working with deformable objects is how to bring the object into a known configuration from an arbitrary initial state.

The pipeline for folding clothes followed throughout robotic literature is based on how it is performed by humans. This allows these tasks to be executed in the same environments as humans, aiming towards a fully-automated or collaborative fashion. Garments are extracted from the washing or drying machine forming a pile, and an iterative process begins. First, an individual clothing article is picked from the pile. Then, the garment is extended in the air or assisted by a flat surface, during which unfolding and wrinkle removing procedures may be performed for aiding the posterior classification and manipulation of the garment. A classification procedure is applied to fit clothing article within a certain garment category. Finally, a standard manipulation sequence specific to its category is applied to fold the garment for storage. This iterative process is repeated until there are no clothes left on the pile.

All of the authors are members of the Robotics Lab research group within the Department of Systems Engineering and Automation, Universidad Carlos III de Madrid (UC3M). destevez@ing.uc3m.es

Extensive work can be found in literature about automated clothes folding once the garment category has been identified, as well as for modeling the garment for fold/wrinkle removal or selecting the most suitable grasping point/strategy. For this reason, our work focuses on how to unfold a clothing article that has been picked up from a pile of clothes and is placed on a flat surface, as seen in Fig. 1.

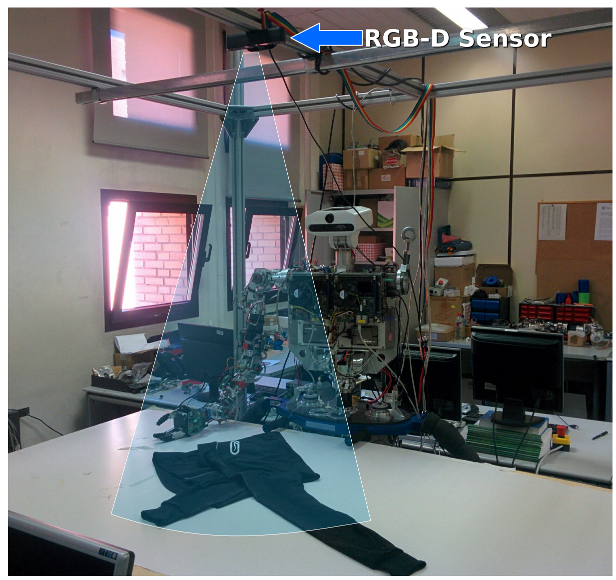

Fig. 1. Humanoid robotics clothes folding scenario. The clothes are placed on a flat white table while a RGB-D sensor is positioned on the top.

The main contribution of our work is an algorithm that can estimate the grasping and release points for a deformable object so that a manipulator robot can iteratively unfold a garment for determining its garment category and the folding sequence to apply. This algorithm presents the following advantages:

- It provides a general method of detecting folds in deformable objects without a prior model of the garment to be unfolded.
- It estimates the best position of the grasping point, direction of movement, and release point in order to unfold the detected fold.
- As it uses depth image information, it is independent of the color or patterns present in the garment, except for background extraction.

This system has been tested with thick pieces of cloth such as towels and blankets. These results can be extrapolated to thinner garments provided the depth sensor provides sufficient resolution.

II. STATE OF THE ART

A significant amount of work conducted in this topic has been focused on modeling the different garment categories for both unfolded, extended garments and for grasped garments. The computer vision community has contributed with extensive work on the specifics of clothes modeling [1].

Kita et al. propose a method that uses a deformable model to calculate the state of hanging clothes based on 3D observed data [2], [3]. This calculation is performed by generating a set of candidate shapes predicted by physical simulations of hanging clothes and later comparison of them with the observed data. To fit the observed 3D data better, each generated shape is further deformed and the shape that is more consistent with the observed data is selected.

Miller et al. present an approach to modeling the clothes when already spread out on a flat surface in [4]. A series of parameterized shape models are proposed, each clothing category having its own model. Garment variability is solved through variation of those parameters. Once the garment has been modeled with their method, a preprogrammed folding sequence can be performed.

A method for classifying and estimating the poses of deformable objects is presented in [5]. This method consists in creating a training set of deformable objects by off-line simulation of different garments, extracting depth images from different points of view. Then, a codebook is built for a set of different poses of each deformable object by extracting features from the dataset and applying sparse coding and dictionary learning. With this codebook, classifying deformable objects on different categories and estimating their current pose is possible, for later regrasping or folding the garment.

The previous method was improved in [6], by extracting the features directly from the 3D data, dividing the hanging garment in different cells via layers, rings and sectors of the bounding cylinder. Each of the sectors becomes a binary feature, using the Signed Distance Function to check if the cell is inside the voxel where the center of the cell belongs, and is then arranged in a feature vector. A Hamming distance, whose weights are learned from the simulated dataset merged with some models reconstructed from real word Kinect point clouds, is used to estimate the object category and pose given an input reconstructed mesh model.

Based on the previous recognition algorithm, Li presents in [7] a method for unfolding deformable objects with a bi-manipulator robot. With this method, the robot is capable of taking a clothing article from an unknown state to a known state by iterative regrasping, detecting the most suitable grasping points in each state to achieve its goal. For locating the most suitable grasping points, the 3D point cloud obtained by the robot is matched to the mesh model, that incorporates the information about the best regions to grasp in order to unfold the garment.

The method introduced by Cusumano-Towner et al. in [8] allows a bi-manipulator robot to identify a clothing article, estimate its current state and achieve a desired configuration, generalizing to previously unseen garments. For that purpose, the robot uses a Hidden Markov Model (HMM) throughout a sequence of manipulations and observations, in conjunction with a relaxation of a strain-limiting finite element model for cloth simulation that can be solved via convex optimization.

Clothing article manipulation is another field in which extensive work has been done. Osawa et al. propose in [9] a method to unfold garments in order to classify them. It consists in alternatively regrasping clothing and expanding them using a two-arms manipulator. The garment is grasped with one arm and the lowest point is located by rotating the piece of clothing, which is used as a grasping point for the other arm. If the garment has any fold when extended, it is placed over a flat surface to repeat this process until the the garment is fully spread out.

To detect the best grasping points for a clothing article, Ramisa [10] performs the identification in a single step, even with highly wrinkled clothes. This detector is based in a Bag of Features detector, using as input a combination of appearance and 3D geometric features.

CloPeMa[1] is a recent EU-FP7 research project (2012-2015) whose objective is to advance the state of the art in perception and manipulation of fabric, textiles and garments. As part of the CloPeMa project, a method to detect single folds has been presented by Mariolis et al. in [11], [12]. In order to detect such folds, a database of unfolded clothes templates is built in the first place. These templates are later used to perform a shape matching between the folded garment shape, obtained by the camera, and the unfolded garment model. This process is iterative, and the initial results are feedbacked to adapt the model for a better fit. Stria et al. propose in [13], [14] a polygon-based model for clothes configuration recognition using the estimated position of the most important landmarks in the clothing article. Once identified, these landmarks can be used for automated folding using a robotic manipulator. The clothes contour is extracted from a RGB image and processed using a modified grabcut algorithm and dynamic programming methods are used to fit it to the polygonal model. Doumanoglou et al. follow in [15] an approach based on Active Random Forests to recognize clothing articles from depth images. This classifier allows the robot to perform actions to collect extra information in order to disambiguate the current hypotheses, such as changing the viewpoint. In [16] they extend this approach to detect the optimal grasping points to unfold the garment.

The most similar work we can find in the related literature is the method for unfolding clothes presented by Willimon et al. in [17]. Their method, which also focuses in clothes unfolding prior to automated folding, use several features obtained from a depth image, such as peak regions and corners location, to determine the location and orientation most suitable for interaction with the garment. Two main steps are performed: first, the clothing article is flattened using RGB information from the camera and, then, depth

[1] http://www.clopema.eu/

information is used to extract the features used to estimate how to unfold the garment.

III. ALGORITHM

In this paper we present a garment-agnostic algorithm to unfold clothes using 3D sensor data. Our work differs with most of the state of the art in the fact that garment models are not used. Our approach uses information extracted from a depth map to detect folds in a garment. The most suitable grasping position, unfolding direction, and release point are computed.

It is assumed that a clothing article has already been separated from the rest of the clothes to fold and placed on a flat surface. The garment could have been placed on that surface either by a robot or by a human coworker, allowing a collaborative folding pipeline in which a human and a robot can perform different parts of the folding process. As our algorithm is not based on a geometrical model of the garment to be unfolded, it is general enough to be used with any category of garment, from towels and blankets to trousers or shirts, and with any number of folds.

The algorithm can be divided into three main stages. First, the background is extracted from the image and the contour of the garment is approximated. Then, a height analysis is performed in order to estimate overlapping regions. Finally, the information obtained in the previous analysis is used to determine if there are any folds present and how to interact with the garment to unfold it. The algorithm has been open sourced and is available online[2].

A. Background Extraction and Contour Detection

Prior to any cloth analysis, a background extraction has to be performed in order to remove all information not related to the garment. Based on the assumption that the clothing article is placed on top of a flat white surface, the garment is segmented by a threshold operation on the saturation and value channels of the HSV space. Another potential approach is to remove the horizontal plane of the table. This approach can be further improved by combining the results obtained with the depth-based segmentation with a color segmentation, in case the garment color and the background are dissimilar enough.

Once the background is extracted, the contour of the garment is extracted and approximated to a polygon using the Ramer-Douglas-Peucker algorithm. Each of the segments obtained is a candidate for being a fold axis. The midpoints of those fold axis candidates are calculated to be used in the following steps.

B. Heightmap Analysis

Once the garment has been identified, garment regions with similar height points must be labeled. For this task we apply the watershed algorithm[3] to the depth image.

[2]https://github.com/roboticslab-uc3m/textiles
[3]http://scikit-image.org/docs/dev/auto_examples/plot_watershed.html

The watershed algorithm treats pixels values, in a grayscale image, as heights. The algorithm floods basins (low regions), until basins attributed to different regions meet on watershed lines.

The image returned by the watershed algorithm is labeled by regions of similar heights, Fig. 2. By using only depth information, as opposed to using RGB images, makes our algorithm independent of the colors and patterns present in the garments.

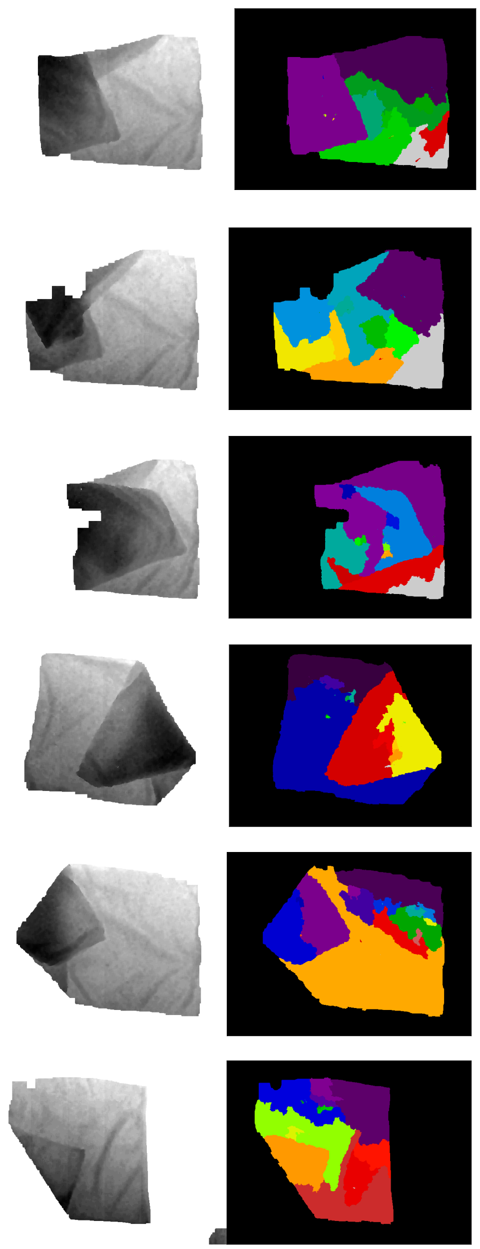

Fig. 2. On the left side, the grayscale images are shown. The grey level is related to the height of the point as detected by the RGB-D sensor. On the right side, the labeled image returned by watershed algorithm is presented, where each color represents a region of similar height.

On the labeled watershed image, we calculate the average height value of each region, and assign this average value to all points in the region. The region with the highest average height is selected for further analysis.

C. Path Detection

The next step, after having labeled the similar-height regions and found the highest region, is to find the unfold direction. The assumption made here is that the fold has at least one contour edge which is also a border of the garment.

A set of candidate paths is generated, all of them starting at the centroid of the highest region, and ending in each contour segment midpoint. Each candidate path is analyzed and assigned a *bumpiness* value B. This value is calculated by penalizing the changes in the height of the path, Eq. (1).

$$B = \sum_{i=1}^{n} |\text{path}(i) - \text{path}(i-1)| \tag{1}$$

Where n is the number of points composing the path. The path with the lowest bumpiness value, which corresponds to the path with the less and smallest height changes, is selected as the unfold direction, Fig. 3.

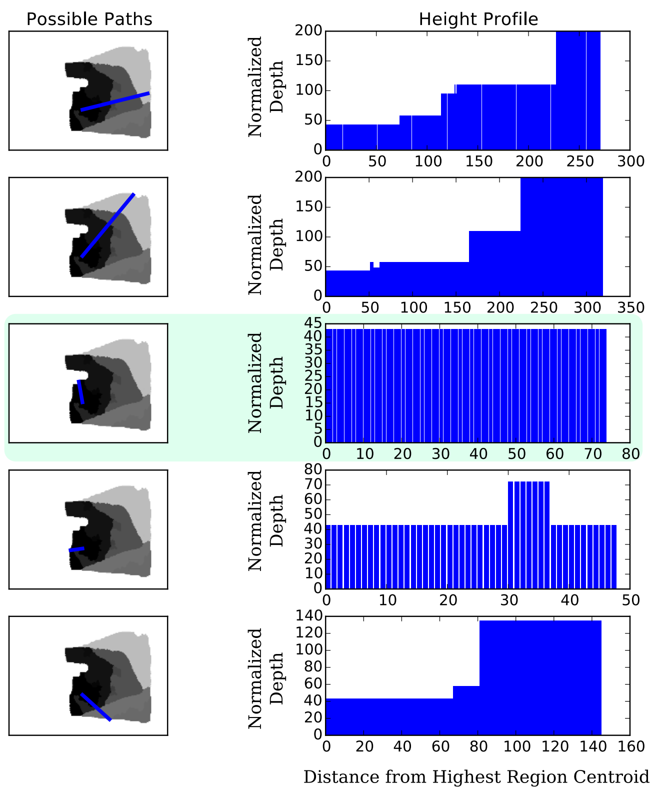

Fig. 3. On the left side, the candidate paths are shown. On the right side, the height profile of each path is shown. Notice that the depth sensor computes the distance to the object from itself, so that a low value in the bar plot means a closer object to the sensor, so it is a region with more height with respect to the table.

D. Pick and Place Points

Up to here, we have detected the most promising direction to unfold. Finally, we need to define is the exact point where the robot has to pick the garment, and the place point.

For this, we extend the previously selected direction line. The garment pick point is chosen at the intersection of the direction line with the inner region border. To find this point, we compute the intersection between the line and the region contour, and select the point furthest to the garment border.

On the other side, the place point is selected by computing the symmetric point of the pick point with respect to the garment edge intersection point. The unfold directions, departing from the pick point and arriving at the place point of a set of clothes, is shown in Fig. 4.

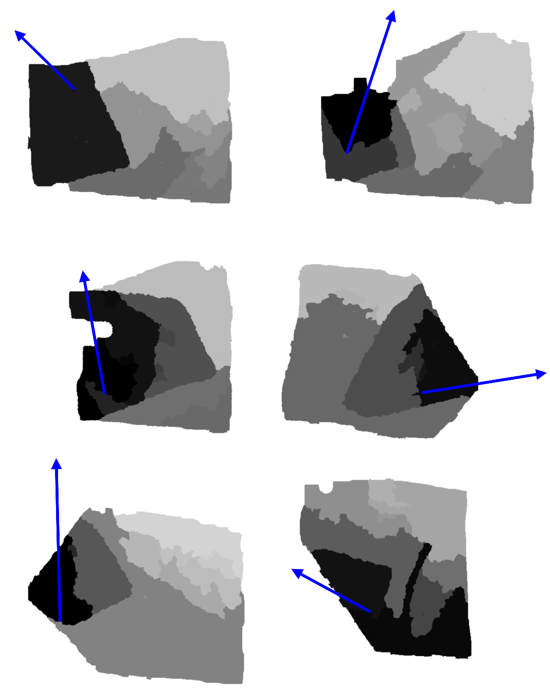

Fig. 4. Final directions calculated for each garment provided to the system. Each arrow departs where the robot should pick the fold and arrives where it should be placed.

IV. EXPERIMENTAL RESULTS

The final experiments using the current algorithm have been performed using an ASUS Xtion PRO LIVE at 640x480 RGB and depth streams (30 fps). To simplify the image analysis, the sensor is placed on top of the working surface providing a bird's eye view over the cloth folding environment, with its image plane almost parallel to the working surface.

In the present work we are using thick blankets and towels, to cope with the limitations of the resolution of the depth sensor. A set of six samples with one or two folds were presented to the system for analysis. The results are shown in Fig. 4. These results show that the algorithm generates acceptable unfolding directions.

According to the results, an interpolation between the best two directions could result in a better solution. The best directions for the set of clothes are shown in Fig. 5.

V. CONCLUSIONS AND FUTURE WORK

In this paper we have presented a model-free garment-agnostic process to unfold clothes. Results show that our approach is promising to be included in a complete pipeline

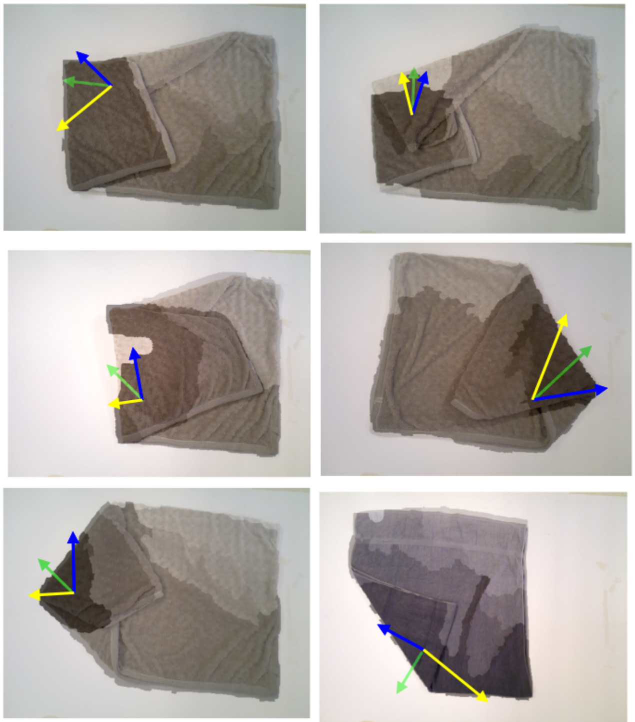

Fig. 5. Best directions calculated for each garment provided to the system. The direction with the smallest bumpiness value is shown in blue. The second best direction is shown in yellow. The bisector is shown in green. The arrows are for demonstrative purpose only, and their starting and ending point do not represent the pick and place point. The watershed computed regions are additionally overlaid upon the original image.

of clothes folding. The main contribution of our work is its analysis of the garment not dependent of a prior model. The process core resides in the segmentation of the grayscale image representing the heights.

Though the results are acceptable, two aspects of the algorithm can be improved. Contour detection currently trims some areas of the garment, which at times results in an incorrect polygon approximation that affects the functionality of the system. The garment pick point, the unfolding direction, and the place point have been chosen in a somewhat arbitrary manner. While the selected points may serve as a rule-of-thumb for robotic system developers, further experiments with robotic systems are required to determine whether a better pick and place strategy exists. For instance, the highest garment point or highest region centroid could be used as alternatives to the current pick point.

ACKNOWLEDGMENT

This work was supported by RoboCity2030-III-CM project (S2013/MIT-2748), funded by Programas de Actividades I+D in Comunidad de Madrid and EU.

REFERENCES

[1] C. Qian, Y. Yan, and S. Ai-fang, "Review of cloth modeling," in *Education Technology and Training, 2008. and 2008 International Workshop on Geoscience and Remote Sensing. ETT and GRS 2008. International Workshop on*, vol. 1. IEEE, 2008, pp. 238–241.

[2] Y. Kita, F. Saito, and N. Kita, "A deformable model driven visual method for handling clothes," in *Robotics and Automation, 2004. Proceedings. ICRA'04. 2004 IEEE International Conference on*, vol. 4. IEEE, 2004, pp. 3889–3895.

[3] Y. Kita, T. Ueshiba, E. S. Neo, and N. Kita, "Clothes state recognition using 3d observed data," in *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on*. IEEE, 2009, pp. 1220–1225.

[4] S. Miller, M. Fritz, T. Darrell, and P. Abbeel, "Parametrized shape models for clothing," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*. IEEE, 2011, pp. 4861–4868.

[5] Y. Li, C.-F. Chen, and P. K. Allen, "Recognition of deformable object category and pose," in *Robotics and Automation (ICRA), 2014 IEEE International Conference on*. IEEE, 2014, pp. 5558–5564.

[6] 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 *Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on*. IEEE, 2014, pp. 1046–1052.

[7] Y. Li, D. Xu, Y. Yue, Y. Wang, S.-F. Chang, E. Grinspun, and P. K. Allen, "Regrasping and unfolding of garments using predictive thin shell modeling," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2015.

[8] M. Cusumano-Towner, A. Singh, S. Miller, J. F. O'Brien, and P. Abbeel, "Bringing clothing into desired configurations with limited perception," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*. IEEE, 2011, pp. 3893–3900.

[9] 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.

[10] A. Ramisa, G. Alenya, F. Moreno-Noguer, and C. Torras, "Using depth and appearance features for informed robot grasping of highly wrinkled clothes," in *Robotics and Automation (ICRA), 2012 IEEE International Conference on*. IEEE, 2012, pp. 1703–1708.

[11] I. Mariolis and S. Malassiotis, "Matching folded garments to unfolded templates using robust shape analysis techniques," in *Computer Analysis of Images and Patterns*. Springer, 2013, pp. 193–200.

[12] ——, "Modelling folded garments by fitting foldable templates," *Machine Vision and Applications*, vol. 26, no. 4, pp. 549–560, 2015.

[13] J. Stria, D. Průša, and V. Hlaváč, "Polygonal models for clothing," in *Advances in Autonomous Robotics Systems*. Springer, 2014, pp. 173–184.

[14] J. Stria, D. Průša, V. Hlaváč, L. Wagner, V. Petrik, P. Krsek, and V. Smutny, "Garment perception and its folding using a dual-arm robot," in *Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on*. IEEE, 2014, pp. 61–67.

[15] A. Doumanoglou, T.-K. Kim, X. Zhao, and S. Malassiotis, "Active random forests: An application to autonomous unfolding of clothes," in *Computer Vision–ECCV 2014*. Springer, 2014, pp. 644–658.

[16] 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 *Robotics and Automation (ICRA), 2014 IEEE International Conference on*. IEEE, 2014, pp. 987–993.

[17] B. Willimon, S. Birchfield, and I. D. Walker, "Model for unfolding laundry using interactive perception." in *Intelligent Robots and Systems (IROS 2011), 2011 IEEE/RSJ International Conference on*, 2011, pp. 4871–4876.