# Determining Where to Grasp Cloth Using Depth Information

Arnau RAMISA [a,1], Guillem ALENYA [a], Francesc MORENO-NOGUER [a] and
Carme TORRAS [a]

[a] *Institut de Robòtica i Informàtica Industrial, CSIC-UPC, Barcelona*

**Abstract.** In this paper we address the problem of finding an initial good grasping
point for the robotic manipulation of textile objects lying on a flat surface. Given as
input a point cloud of the cloth acquired with a 3D camera, we propose choosing
as grasping points those that maximize a new measure of wrinkledness, computed
from the distribution of normal directions over local neighborhoods. Real grasping
experiments using a robotic arm are performed, showing that the proposed measure
leads to promising results.

**Keywords.** Deformable Objects, Computer Vision, Grasping point selection

## Introduction

Recently the problem of grasping and folding clothes with a robotic arm has attracted
much attention [1,2,3,4,5,6]. Its application ranges from automating industrial clean-
ing facilities to domestic service robots. There exist works devoted to determining the
best/optimal grasping point for a particular purpose (e.g. folding) once the cloth is held
by a robotic hand. However, most of the research in this area has been carried out in
controlled environments and simple heuristics have sufficed.

In [3] the authors designed a cloth grasping-point selection system to autonomously
take elements form a pile of washed towels and fold and stack them ready for storage
using a Willow Garage PR2 robot. The main contribution of the paper in terms of grasp-
ing is being able to detect a corner when the towel is already grasped by one of the robot
arms, and the initial pick-up is done by selecting the central point of the cloth, detected
through background segmentation and stereo. The performance of this initial grasp was
not reported in the paper.

In [4] the authors describe a complete system, designed for the PR2 robot, for laun-
dry folding. The system can start from a pile of clothes, pick up and identify one of
them and bring it into a desired configuration; and repeat the procedure until no more
clothes are left. Two HMM are used for the two tasks, identification and manipulation.
The cloth is initially grasped by one edge. This kind of grasping with PR2 manipulators
is only possible because the surface in which the cloth lies is made of a soft material that
deforms under the robotic hand.

A complete system for retrieving one by one all elements of a laundry basket or pile, classifying and then folding them is proposed in [6]. In this approach, the topmost element of the pile is found using stereo vision, and its geometric center is used as grasping point. The grasping operation is repeated as many times as necessary to assure a correct grasp.

Foresti and Pellegrino presented a vision-based system to detect grasping points for furs in an industrial environment [7]. Their system used a hierarchy of Self-Organizing Maps to segment the image into fur and non-fur areas, and then analyzed the detected fur blobs to determine the best picking point for the robotic arm.

Taking a more sophisticated approach, in [5] the authors propose a method for estimating the pose of cloth through parametrized shape models, specific for each category of clothing (e.g. t-shirt, pants, towel). These models were fit using an energy optimization approach, and then used to classify the cloth item type prior to folding it with an open loop movement sequence. Finding an accurate cloth item pose model would definitely help in the task of initial grasping point selection; however, fitting the model to a cloth in each image acquired by the robot can take up to 2.5 minutes when running in a multi-core laptop computer, which may be too expensive for practical applications. Furthermore, to fit the required models the cloth has to be presented to the system in a canonical form, only possible thanks to a previous manipulation step.

In this paper, our purpose is to investigate what constitutes a good initial grasping point for a piece of cloth lying on a flat surface in an arbitrary configuration. We propose a new "wrinkledness" measure based on range information that can be used to determine the most easily graspable point at an affordable computational cost.

Recently, a method related to ours that detects cloth objects using 2D images has been proposed [2]. A cloth detection method in a domestic or office environment based on wrinkle features is presented. The wrinkle features were found by analyzing the response of Gabor filters with a Support Vector Machine (SVM) trained with manually annotated images of wrinkled cloth. Finally, graph cuts were used to segment the cloth pieces from the background. The proposed method was applied in a cloth pick-up task using a mobile robot, which selected its grasping target as the wrinkle with maximal 3D volume according to stereo camera measurements.

Although this approach is very appealing, it differs from ours in a number of points: First, although our approach could be used for the same purposes, [2] tackles the more comprehensive task of detecting clothes in an unprepared environment, and not the identification of optimal grasping points. Second, we directly use 3D information obtained from an economic sensor (therefore avoiding the expensive data collection and manual annotation step for training the SVM), and thus our procedure is not vulnerable to any learning error.

In order to acquire the depth information, we rely on a Kinect 3D camera[2], which is a very affordable device to obtain simultaneously depth and texture maps in an indoor environment, and is becoming of widespread use in the robotics and computer vision communities, directly competing with the much more expensive time-of-flight PMD cameras [8]. Kinect uses an infrared structured light emitter to project a pattern into the scene and a camera to acquire the image of the pattern, then depth is computed by means of structured light algorithms. Additionally, among others sensors, the Kinect integrates a high resolution color camera.

---

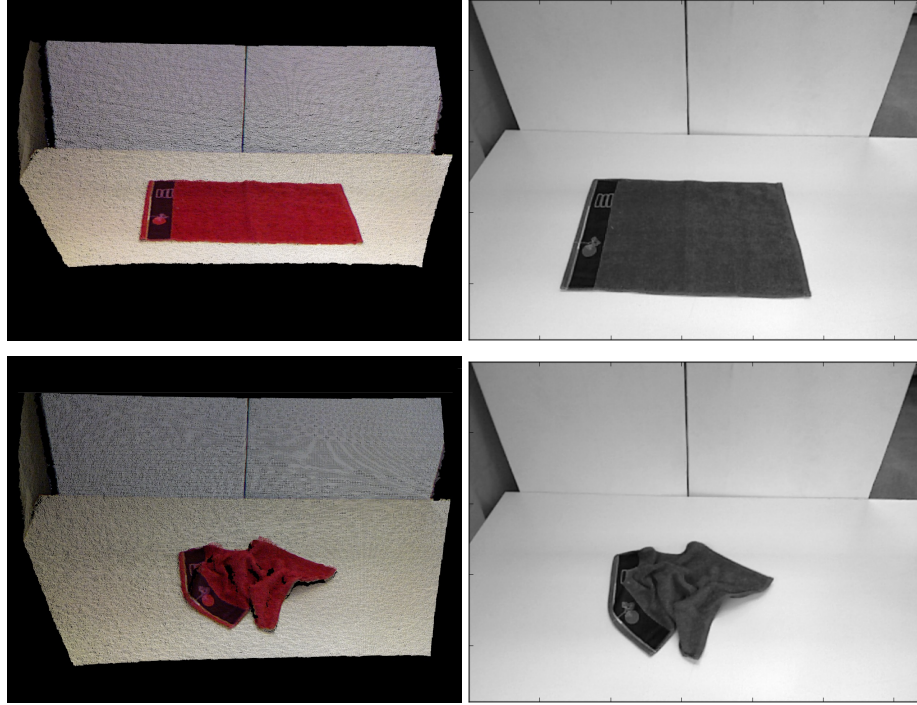[2]Developed by Prime Sense `http://www.primesense.com/`

**Figure 1.** Textured point clouds (left) and images (right) acquired with the Kinect 3D camera. As can be seen, some points are missing due to lack of resolution and occlusions in the wrinkled point cloud.

Kinect was developed with the idea of robust interactive human body tracking and great efforts have been made in that direction [9]. After the Kinect protocol was hacked, the community rapidly started to use it, first with the same idea of human interaction and afterward in other areas, like robot navigation. Later, the official library was made public through the OpenNi organization.

Some pictures and point clouds of a towel acquired with the Kinect 3D camera are shown in Figure 1. As can be seen, the resolution is not sufficient to perceive small details, and some holes without depth information occur.

## 1. Graspable Point Detection

In this section we first present in detail the problem that we are addressing and the dataset collected to validate the method. Next, we describe the proposed algorithm to determine a graspable point given an input image and a depth map of a cloth.

Our initial assumption is that a good grasping point for a textile object lying on a table is one where the cloth defines ridges or other 3D structures, i.e. where there are wrinkles. The justification of this assumption comes from the nature of the grasping mechanism, which in our case has three fingers, with a total of four degrees of freedom. Lacking the precision of movement, flexibility and the small(er) size of human hands

(which can pick up cloth objects from the edges), the best point for a grasp with this type of hand is a pyramidal or conic-like shape, such as the one produced by wrinkles.

One common heuristic or workaround used by works addressing textile grasping such as [3,6] is to select as grasping point the highest one in the 3D point cloud of the cloth object. In practice, the highest point usually coincides with a wrinkle, and thus it is in agreement with the assumption stated above. However, the highest point is not necessarily a good grasping point in all situations, as we show in the experimental section of this paper (see Section 2). The objective of this work is to explore how to characterize the "wrinkledness" of a cloth object, and to find out its advantages and drawbacks for detecting good grasping points. Ultimately, this cue can be used in the design of a method able to determine the best grasping point in a robust way.

### 1.1. Proposed Wrinkledness Measure

We have developed a measure of the "wrinkledness" in a point taking into account the depth information of its neighborhood. This measure is computed using a local descriptor based on the surface normals of a 3D point cloud. In particular, we use the *inclination* and *azimuth* angles in the spherical coordinates representation of the normal vectors:

$$(\phi, \theta) = \left( arccos\left(\frac{z}{r}\right), arctan\left(\frac{y}{x}\right) \right) \tag{1}$$

where $\phi$ is the inclination and $\theta$ is the azimuth, $(x, y, z)$ are the 3D point coordinates, and $r$ is the radius in spherical, defined as:

$$r = \sqrt{x^2 + y^2 + z^2}. \tag{2}$$

Next, we model the distribution of the inclination and azimuth values in a local region around each point. Although it would be very interesting to introduce spatial subdivisions on the local region and model the distribution of the angles locally in each subdivision, for our current purposes we found it was not necessary and therefore we left it as future work.

We evaluated two possibilities for modeling the spherical coordinate angles distribution: two sixty-four-bin histograms, one for each angle, and a single two-dimensional histogram with $64 \times 64$ bins that considers both angles jointly.

A beneficial side effect of this process is that occluded regions and areas where the Kinect was not able to estimate the depth are naturally interpolated using the information provided by their neighbors, which reduces the sparsity of the point cloud.

From this model of the local distribution of normal angles in spherical coordinates, we seek to estimate the "wrinkledness" of a point. This can be intuitively done by looking at the spread of the angle histogram: the more different orientations the surface takes, the more likely that it is a highly wrinkled area. Although standard deviation is probably the first measure of spread that comes to mind, it is not a good choice, since a strongly bimodal distribution can have a large standard deviation while having low spread. A better choice is entropy, which does not suffer from this drawback:

$$H(X) = -\sum_{i=1}^{n} p(x_i)\, log\, p(x_i) \tag{3}$$

where $X$ is the $n$-bin angle orientation histogram, and $x_i$ is the $i$th bin.

Entropy measures how much information exists in a message or distribution, or alternatively, how "predictable" it is. In our context, it directly tells us the amount of support of the distribution concentrated in high probability peaks or, equivalently, how much of the surrounding area of the point has normals aligned in the same orientation i.e. a flat surface or a combination of a few flat surfaces. Entropy has the additional advantage of not assuming an unimodal distribution like standard deviation or high-order moments such as kurtosis and skewness.

In Figure 2 "wrinkledness" maps using the proposed entropy-based measure can be seen for the 1D and the 2D histogram representations. It can be observed that the 2D measure produces slightly clearer maps with a longer range of activation levels, faithfully reflecting the wrinkled areas of the towel. Of course it is possible to vary the support region of the descriptors (or downsample the point cloud). This would allow to obtain a smoother result which only reflects a few "global" maxima with large support areas, and peaky "wrinkledness" maps capturing all the local maxima when a small support region is used. Figure 3 shows the response obtained with different support regions. One important limitation of this approach is that concave areas of the image get a high activation level while not being a good grasping point. Yet, it is possible to compute a concavity measure and use it to re-weight the "wrinkledness" map. Finally, the peaks of the map can be used as candidate grasping points.

## 2. Experiments

We tested our proposed "wrinkledness" measure in real grasping experiments. Our experimental setup consists of a robotic hand with three fingers installed in front of a flat table of uniform color, in which the cloth object was positioned. We mounted a Kinect 3D camera having a zenital view of the table and providing the information used to select the grasping point. The piece of cloth used in the experiments is the previously seen small red towel which, for the purpose of evaluating our method, is segmented from the background using a simple color thresholding procedure, as done in similar works.

Five experiments were performed with different initial configurations of the towel. In all cases, the 2D histogram with a square support region with a side of 33 pixels was used to generate the "wrinkledness" map after segmenting the towel from the table, and the point with the highest activation was selected as the grasping point. Next the robotic arm was moved to the point, and a grasp attempt was performed. Please note that we are not claiming that the point with highest activation in the map is necessarily the best grasping point. However here we used this simple heuristic with very good results. Four out of five tests ended with a successful grasp. Figure 4 shows the images and "wrinkledness" maps used to decide the grasping point, and a photo of the robotic arm holding the towel for those tests that were successful. In each successive test the towel was positioned in an increasingly difficult configuration.

- In the first test, there was one clear peak in the center of the elsewhere flat towel, which was correctly detected and ended with a successful grasp.
- For the second test, the towel contained a similar wrinkle, but this time at the corner, making the task more difficult (smaller support region, less clear wrinkle).
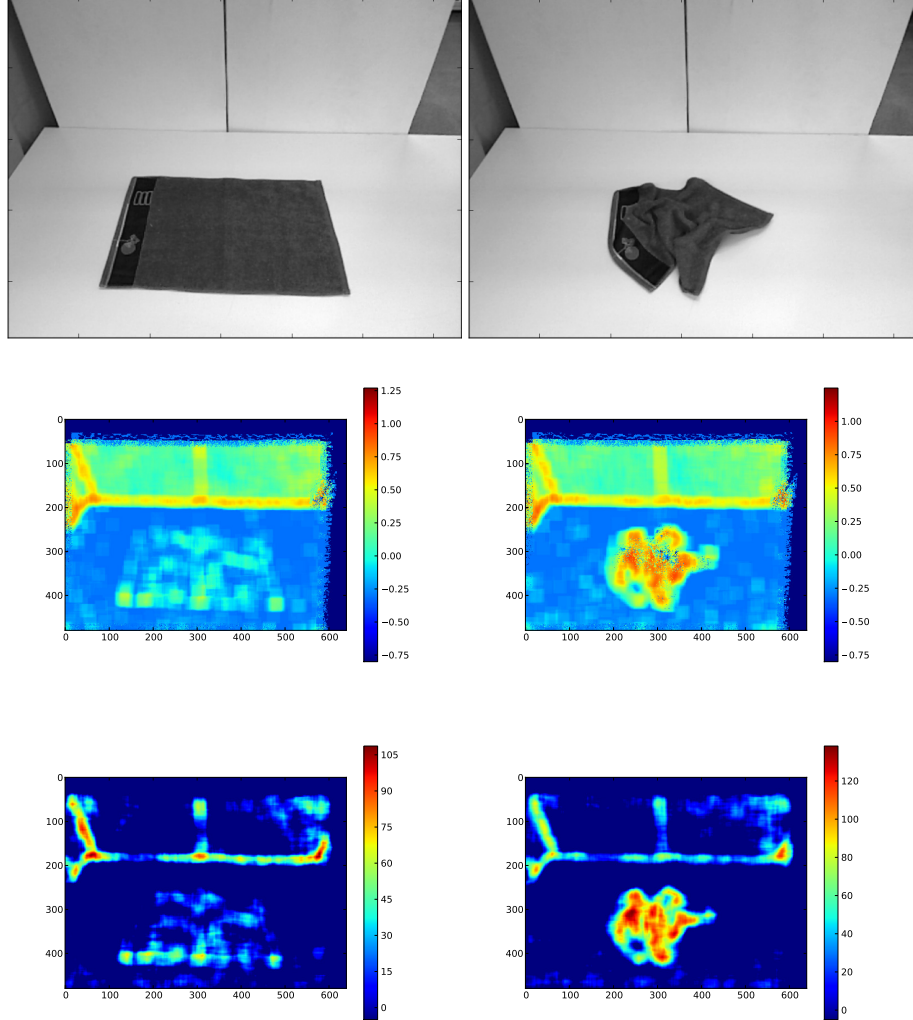
**Figure 2.** Response of the proposed "wrinkledness" measure. First row: original images, second row: "wrinkledness" map for the 1D histograms measure, third row: "wrinkledness" map for the 2D histograms measure. The considered local area around each point is of 33 pixels. Please note the difference in the colorbar range.

However, the method was able to find a good grasping point and the experiment was successful.

- The third test exemplifies our motivation to characterize good grasping points. In this test, the towel presents a configuration in which the highest point is not good for grasping (it is an almost flat surface), but a lower point is. The presented wrinkledness measure finds the right (lower) point and the grasp ends successfully.
- In the fourth test, most of the area of the towel is concealed by a fold, and only two small wrinkles are present in the uncovered area. Our method does not find the best grasping point from a human perspective, but the selected point is not
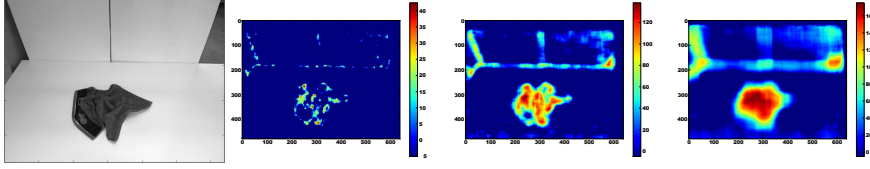
**Figure 3.** Response obtained with different support regions for the local descriptor. From left to right: Original image, "wrinkledness" maps with support region $15 \times 15$, $33 \times 33$ and $65 \times 65$.

completely bad either and the grasp ends successfully as well. The confusion of the method in this case is due to the merging of two different layers of cloth in the same local region. A drawback of selecting this kind of points is that it can lead to grasping the towel from two separate points at a time, making it more difficult to succeed in the subsequent tasks (e.g. grasping the towel from the edges for folding) to succeed. Introducing a continuity-enforcing measure to the local region would help to prevent this situation.

- Finally, in the last experiment, the robotic hand did not find any graspable surface in the selected point. The failure is due to a concavity being detected as the point with highest entropy in the orientation of the normals. As mentioned earlier, for the proposed measure to be more robust, concavities should be detected and down-weighted in the "wrinkledness" map.

## 3. Conclusions

In this paper we have presented some preliminary work towards finding a good measure of "graspability" for cloth objects lying on a flat surface. This is an important aspect for making robots fully autonomous in unprepared environments; in contrast, related literature so far relied on simple heuristics, that worked in controlled settings.

Our proposed measure is computed from point clouds acquired with a Kinect 3D camera, and uses entropy in the normal vector orientation distribution around a given point as an indicator of wrinkledness. Wrinkled areas constitute, in general, good graspable points.

Although not addressed here, this measure can be later combined with other cues to make it more robust. In particular, concave areas pose a problem since they yield a high score with our measure, but they are in general not good for grasping.

We performed real grasping experiments with the proposed measure. For them, a towel placed on a table and observed by a Kinect 3D camera from a zenital position, was grasped from the point of maximum entropy in the normal directions of the towel surface. The towel was previously segmented from the table using color. Four out of five tests ended in a successful grasp, showing that the presented measure is suitable to be used for identifying good grasping points.

In terms of computational cost, our non-optimized implementation is able to compute a dense "wrinkledness" map in a few seconds. Moreover, it is possible to make the grid more sparse without affecting too much the quality of the results (e.g. computing
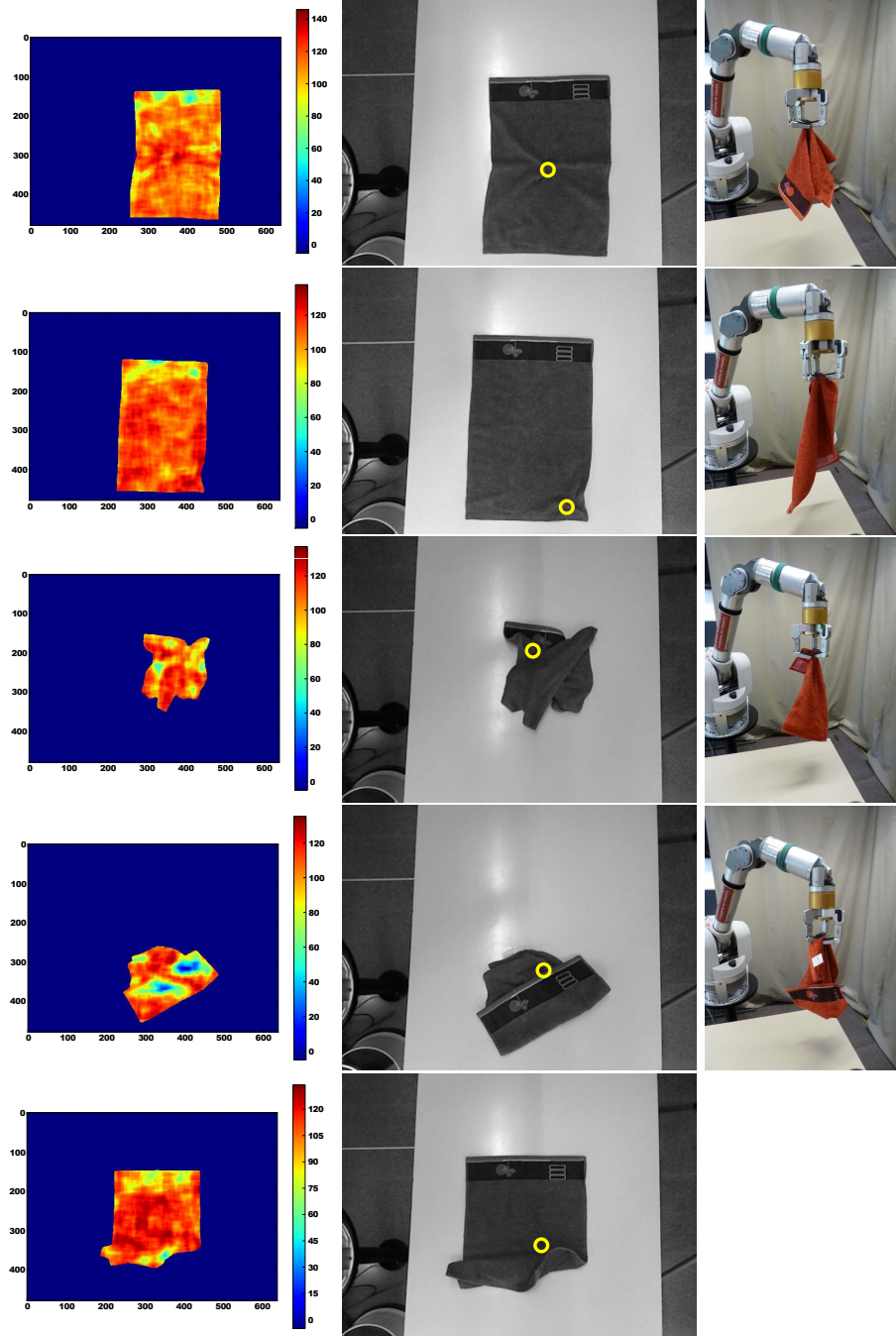
**Figure 4.** Details of the five experiments conducted with a robotic arm (one per row). For each experiment the following are shown (in order): the segmented "wrinkledness" map of the towel, the selected grasping point, and a picture of the robotic hand with the grasped towel, if successful.

the entropy at every three pixels instead of at every pixel) would significantly reduce the computation time.

Future work includes more thoroughly evaluating the proposed measure to identify its weaknesses, and researching other cues that can help make it more robust.

One point worth exploring is a concavity detector to avoid selecting points where no graspable surface can be found. While simple algorithms can suffice for this task, they will provide a considerable improvement of the final "graspability" measure we are pursuing.

Another interesting future work would be investigating how to improve the normal orientation descriptor by computing it in subdivisions of the local region around the point, using soft voting to reduce the effect of aliasing occurring at orientation bin boundaries or varying the number of orientation bins. Moreover, taking advantage of a multi-scale representation of the "wrinkledness" map to better characterize optimal grasping point locations would be interesting as future work.

Finally, better grasping points could be found by combining information like point height, total 3D volume, normal orientation or the aforementioned concavity measure with the entropy-based measure proposed in this paper.

## Acknowledgements

## References

[1] F. Osawa, H. Seki, and Y. Kamiya, "Unfolding of massive laundry and classification types by dual manipulator," *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol. 11, no. 5, pp. 457–463, 2007.

[2] K. Yamakazi and M. Inaba, "A cloth detection method based on image wrinkle feature for daily assistive robots," in *IAPR Conf. on Machine Vision Applications*, pp. 366–369, 2009.

[3] 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 *Robotics and Automation (ICRA), 2010 IEEE International Conference on*, pp. 2308–2315, IEEE, 2010.

[4] M. Cusumano-towner, A. Singh, S. Miller, J. F. O. Brien, and P. Abbeel, "Bringing Clothing into Desired Configurations with Limited Perception," in *2011 IEEE International Conference on Robotics and Automation (ICRA)*, (Shangai, China), pp. 3893–3900, 2011.

[5] S. Miller, M. Fritz, T. Darrell, and P. Abbeel, "Parametrized Shape Models for Clothing," in *2011 IEEE International Conference on Robotics and Automation (ICRA)*, (Shangai, China), pp. 4861–4868, 2011.

[6] B. Willimon, S. Birchfield, and I. Walker, "Classification of Clothing using Interactive Perception," in *Proc. IEEE International Conference on Robotics and Automation (ICRA11)*, pp. 1862–868, 2011.

[7] G. Foresti and F. Pellegrino, "Automatic Visual Recognition of Deformable Objects for Grasping and Manipulation," *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)*, vol. 34, pp. 325–333, Aug. 2004.

[8] S. Foix, G. Alenyà, and C. Torras, "Lock-in Time-of-Flight (ToF) cameras: a survey," *IEEE_J_SENSOR*, vol. to appear, 2011.

[9] J. Shotton, A. Fitzgibbon, M. Cook, and T. Sharp, "Real-time human pose recognition in parts from single depth images," in *Computer Vision and Pattern Recognition (CVPR)*, p. to appear, 2011.