Teaching Grasping Points Using Natural Movements

Yalim ISLEYICI [a] and Guillem ALENYÀ [b]
[a] *Universitat Politècnica de Catalunya (UPC)*
[b] *Institut de Robotica i Informàtica Industrial (CSIC-UPC)*

Abstract. Recently, the research on robots performing every-day tasks at home, like to take care of elderly or disabled people, has pursued the problem of the manipulation of everyday objects. Among them, grasping a cloth is one of the most challenging tasks, since the textiles are highly-deformable and it is not straightforward to define a generic grasping point. In this paper, we address this problem by introducing a new robot interaction method that enables unexperienced users to control the robot in a natural way. Given a textile lying on the table the robot proposes a grasping point and the user is able to teach the robot a new grasping point. The data collected using this method is then used for training the system using a Vector Autoregression method, which produces better grasping points allowing potentially better manipulation actions. The experiments demonstrates the validity of the new interaction method and its potential to improve the point-grasping selection algorithm in different configurations of a polo-shirt.

Keywords. Textile grasping, Human robot interaction, Leap Motion, Teleoperation, Reinforcement Learning, Cloth manipulation

1. Introduction

Need for robotic maids to take care of elderly and disabled people caught interest of many researchers. The tasks that are expected from robots vary in complexity. They may be a part of series of tasks such as opening a door [1] or a set of complex tasks, such as cooking [2]. However, one thing that the robotic maids have in common is, the robots and the people should interact while the robots are helping them. There are numerous ways for human robot interaction (HRI). There are ways to manipulate directly by applying force to the arms of the robot such as Mobile Robot (MR) Helper, MR Dancer as explained in [8], using joysticks and using common input devices like a keyboard and a mouse.

The problem with these methods are either robots are operated in close proximity or the used teleoperation device does not represent the robot movements in an easy reference frame for a non-experienced user. As most of the helpers are designed to help the elderly, i.e. non-experienced users, it is crucial that a person without any prior knowledge should be able to operate the robot easily. To address this problem, numerous sensors are used that can recognize the human body parts and their movements and hence enabling gesture based operation. Robots with a Kinect sensor (Microsoft, USA) available, can easily detect a human body and its movements and act accordingly [9,10]. Leap Motion

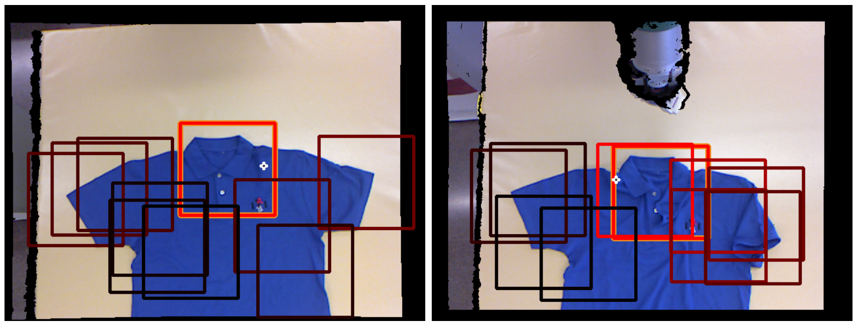

Figure 1. Bounding boxes and FINDDD selected grasping point. The red box contains the collar and FINDDD scores will be calculated inside it.

sensor (Leap Motion Inc., USA) is a new sensor that can recognize hands and track its position alongside the separate positions of fingers using IR sensors and cameras. Having a 150°of detection area and accuracy in sub-millimeters, this device is being used in teleoperation of robots, especially in robot arms [11].

One of the main tasks that a robotic maid should be able to handle is cloth manipulation. As the textile is a deformable object, it is not trivial for a robot to select a grasping point on a cloth that is acceptable to humans, or alternatively, a good grasping point taking into account the task to be performed. There are numerous attempts to solve this problem. For example in [3] the robots folds a cloth from a cluster of shirts. In [4] Fast Integral Normal 3D (FINDDD) descriptors are used to detect the collar on a polo-shirt and how to select a point based on this data is explained. However, as seen in Figure 1, FINDDD is not able to always find a good grasping location that is acceptable to humans. Therefore, the system can benefit of a training case by case by human beings.

In this work, a method to teleoperate the robot in a more natural way for reinforcement learning is proposed. Later its application for training an autoregression model based on FINDDD descriptors will be shown. In the first part, the methodology of teleoperation with the use of Leap Motion sensor is explained. Later, the autoregression method and the results obtained will be shown. In the last section, obtained results will be discussed.

2. Selecting Better Grasping Points

2.1. Human Interference to Introduce a Better Grasping Point

In order to teach a robot to select better grasping points than the previous ones, a human intervention is needed. This intervention might be with direct manipulation or with haptic devices. However in this work, a more natural way will be used. With the help of Leap Motion sensor, which can detect and track hands, user can move a robotic arm in a more natural way and it will improve the sensation of teleoperation. (Step 4 in Figure 2). Moreover, this method allows an unexperienced user to operate the robot with ease.

With the help of Leap Motion sensor, the robot gripper can mimic the position and the orientation of the hand. Moreover, it can mimic the positions of the fingers. With a simple gesture such as finger tapping, the user can command the robot to execute

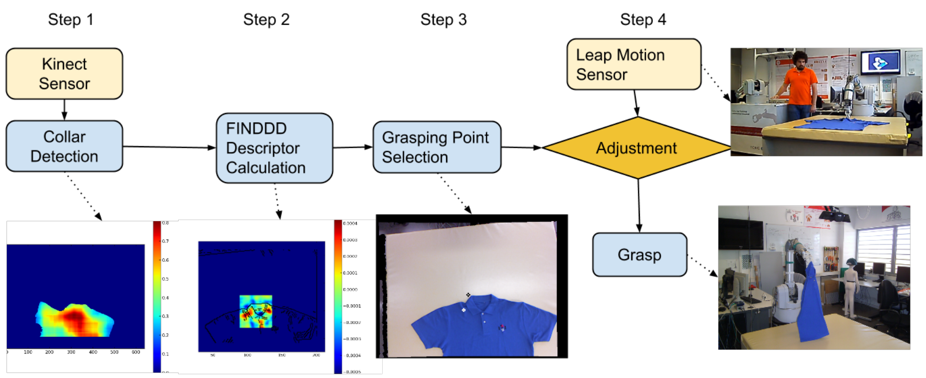

Figure 2. Complete flowchart of the system used.

a task such as grasping. However, the tremor on the hand causes the robot to behave undesirably. In addition, the noisy information from the sensor forces the robot to act beyond its acceleration and torque limits. To avoid these problems, a first order low-pass butterworth filter with cut-off frequency $\omega = 0.5$ is implemented. Its calibration is done such that, the delay caused by the filter is unnoticeable to the user and one can position the robot precisely. To improve the precision, hand movements are scaled down by a factor of 2 before sending the positions to the robot.

2.2. *Collar Detection*

In order to select a good grasping point on a shirt, its collar is to be found. In [6] a method to detect a bounding box that contains the collar based on Bag of Features [5] with combination of appearance and 3D geometry is explained. (Step 1 in Figure 2). However, locating the collar is not enough for selecting a good grasping point. Therefore in [4] FINDDD descriptors which is an improvement to the method in [6] is proposed. The selection method by this algorithm is calculating FINDDD descriptors inside the bounding box that contains the collar and selecting the grasping point as the location corresponding to the maximum of calculated scores. Having more than one local maximum as seen in Figure 3, this method is highly dependent on chance which is clear in Figure 1. Moreover, even though upper-middle part of the collar is mostly selected by humans, there is no local maximum in FINDDD scores around this region. Hence it is impossible for this method to select a grasping point here. However the distribution of FINDDD descriptors is able to indicate the shape of the collar and it will be used for training and testing the system proposed in this work.

2.3. *Training*

To train the system for selecting better grasping points, a mapping with vector autoregression (VAR) method [7] is used. Eventhough VAR model is mostly used in finance and economics with time variant values, its multivariate version can be modeled for this case to find a mapping matrix A to match the features extracted from FINDDD scores Y and their corresponding user selected grasping points X.

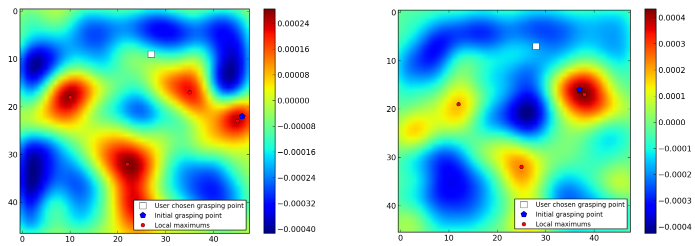

Figure 3. FINDDD Scores

$$\begin{bmatrix} x_{11} & \cdots & x_{1n} \\ x_{21} & \cdots & x_{2n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mn} \end{bmatrix} = \begin{bmatrix} y_{11} & \cdots & y_{1k} \\ y_{21} & \cdots & y_{2k} \\ \vdots & \ddots & \vdots \\ y_{m1} & \cdots & y_{mk} \end{bmatrix} \times \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2n} \\ \vdots & \ddots & \vdots \\ a_{k1} & \cdots & a_{kn} \end{bmatrix} \tag{1}$$

$$X = Y \times A \tag{2}$$

Hence, we can calculate A by,

$$A = ((Y^T \times Y)^{-1} \times Y^T) \times X \tag{3}$$

As explained earlier, it is clear that the grasping point should be moved in the x and y directions in order to achieve grasping from the upper-center part of the collar. It is also necessary to grasp with a correct depth and orientation. For example, grasping deeper than necessary prevents the robot from further manipulations. In addition, the grasping angle θ should be correct in order to enable further manipulations. The complexity of necessities justifies the need of a direct human intervention while the robot is performing the task.

In this work, grasping positions x, y, z and gripper orientation θ is to be found. Therefore, matrix X will have 4 columns formed by these parameters. The values on the Y matrix has to be determined as values that can show the position of the collar. When the FINDDD scores in Figure 3 are analyzed, 4 local maxima with similar positions with respect to collar is observed. Hence, the positions and values of the mentioned points are selected as feature space.

3. Experimental Work and Results

In this section, the experimental work that has been done to test the VAR algorithm and the results obtained from them will be explained.

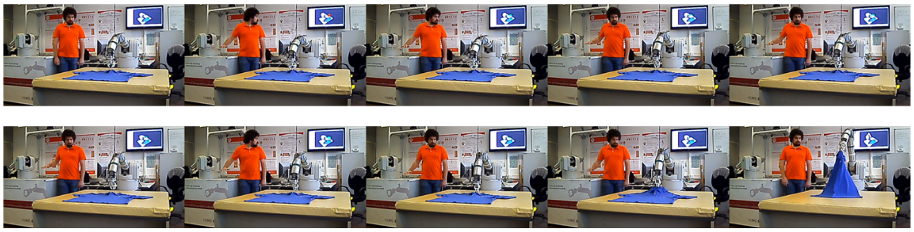

Figure 4. Experimental sequence.

3.1. Experimental Setup

To solve the aforementioned problem, a setup consisting of 7 DoF WAM robot arm (Barret Inc., USA) and for teleoperation Leap Motion sensor are used. The detection of the shirt and the collar is done by a Kinect sensor mounted on top of the table. With the use of Leap Motion sensor, the user is able to control the robot with his natural movements. Moreover, it allows scaling as compared to direct manipulation which allows more precise positioning. This is crucial as the main objective of this work is fine tuning of the grasping positions.

Experimental method is as follows:

1. Polo shirt is placed on the table.
2. With the use of Kinect Sensor, FINDDD scores are calculated.
3. The robot moves to the grasping position selected by the scores calculated in the previous step.
4. The user takes control of the robot and shows a better position and angle for grasping.
5. FINDDD scores and the user selected grasping position are recorded for training purposes.

The entire process can be seen in Figure 4.

3.2. Training

For training a better system, a data set of 50 sample is formed by the methods explained in Section 3.1. The data set is formed by the FINDDD scores for the collar area of the polo shirt and the grasping points determined by the operator. As shown in Figure 3, locations of the first 4 highest valued local maxima with respect to the collar area follows a pattern on the 2D matrix of FINDDD scores. Thus these locations alongside with their values are selected as the feature space.

The feature matrix Y is formed by concatenating the pixel positions of local maxima and their values as seen in Equation (4).

$$Y = \left[\vec{u_{max}}^{T}\ \vec{v_{max}}^{T}\ \vec{z_{max}}^{T}\ \vec{score}_{max}^{T}\right] \tag{4}$$

The outputs matrix X is formed by concatenating the pixel values of user given position on 2D FINDDD scores image u and v which are corresponding to the real world

x and y values, and real world z and θ values as seen in Equation (5). Out of 50 samples, 80% is randomly selected for training and the remaining 20% is selected for testing.

$$X = \left[\vec{u}^T \; \vec{v}^T \; \vec{z}^T \; \vec{\theta}^T\right] \tag{5}$$

Using the Equation (2) we can write,

$$X_{train} = Y_{train} \times A \tag{6}$$

and using the Equation (3) we can calculate the training matrix A.

3.3. Experiments

To evaluate the performance of this algorithm, a simple case of spread open shirt is used. The experiments conducted as explained in Section 3.1 with polo shirt placed on table in a spread open way, as seen in Figure 1. The grasping position for the end effector is selected such that the gripper is perpendicular to the collar, so that it will not cluster the shirt and with just enough depth for a successful grasp.

To evaluate the obtained results from the regression model, the 3D distance from the user selected point to the regression output is checked for each testing sample. However, closer result does not necessarily mean a good result. The output may lay outside of shirt. Therefore, the model output is checked on the real image to see whether the output is actually on the shirt or not.

The error on z direction and on the angle θ is checked separately as their effect is different from the effects of x and y directions and the robot moves to the initial grasping point with constant depth and orientation.

Moreover, the results of the system are tested for the stability regarding the change of button states. The experimental procedure and result evaluation are the same as the previous case.

3.4. Results

The results obtained from the experiments show that using autoregression model in fact moves the grasping point closer to the user requested position. As seen from the Table 1, for all of the 10 testing samples with two buttons closed, the results moved closer to the user selected point and moving the average error from 7.44 cm to 2.28cm. Moreover, when the z and θ differences are checked from Table 2, we can observe that the depth difference average is 3.44 mm and θ error is 0.159 radians which is in an acceptable range.

When we analyze the one button case training from Tables 3 and 4, we can observe that the grasping point does not always moves closer. Moreover, there is one case where the autoregression model is not able to find a solution inside the bounding box. However in average, the 3D distance error decreased from 8.42 cm to 4.53 cm. The error in z direction and θ orientation is 6.86 mm and 0.169 rad respectively which are worse than the initial case but still in an acceptable range.

Regarding the second part, that is to check if the closer grasping point is in fact a better point, the obtained outputs are shown on the real image, as seen in Figure 8 Results

Test data	Initial error (cm)	Error after training (cm)
No: 1	5.58	3.39
No: 2	8.91	1.85
No: 3	7.67	3.73
No: 4	8.02	1.06
No: 5	10.0	1.53
No: 6	6.06	0.06
No: 7	4.59	1.93
No: 8	7.74	0.00
No: 9	7.44	5.50
No: 10	8.46	3.75

Table 1. Position error before and after training for two buttons zipped case

Test data	z direction error (mm)	θ difference (rad)
No: 1	8.91	-0.11
No: 2	-5.60	-0.07
No: 3	1.91	0.10
No: 4	0.39	0.20
No: 5	-5.30	0.05
No: 6	4.11	0.31
No: 7	0.29	0.09
No: 8	1.34	0.20
No: 9	4.06	0.22
No: 10	-2.49	0.24

Table 2. Error on z direction and angle θ for two buttons zipped case.

Test data	Initial error (cm)	Error after training (cm)
No: 1	5.21	6.53
No: 2	7.31	4.45
No: 3	9.10	1.21
No: 4	9.23	1.90
No: 5	8.93	13.5
No: 6	8.53	5.18
No: 7	8.38	Not found
No: 8	7.93	3.68
No: 9	11.55	0.67
No: 10	8.09	3.68

Table 3. Position error before and after training for one button zipped case

on the shirt and closer to the upper-center part of the collar are regarded as good results (Figure 5) and results not on the desired area (Figure 6) or outside of the shirt (Figure 7) are regarded as worse results.

With this method 20 results are controlled and 65% of the results are classifed as good results.

Test data	z direction error (mm)	θ difference (rad)
No: 1	2.79	0.01
No: 2	18.7	0.30
No: 3	5.47	-0.16
No: 4	13.1	-0.21
No: 5	10.1	-0.34
No: 6	4.71	0.36
No: 7	-5.63	0.14
No: 8	2.67	-0.07
No: 9	-2.83	0.02
No: 10	-2.67	-0.08

Table 4. Error on z direction and angle θ for one button zipped case.

(a) Better grasping point after training (b) Worse grasping point - On shirt (c) Worse grasping point - Outside of shirt

Figure 5. Results for two button zipped without rotation (Light cross: initial selected point, Dark cross: Autoregression output)

4. Discussion and Future Work

In this paper we have presented a new interaction method using the novel Leap motion device that allows to non-expert users to interact naturally with a manipulator robot. We have also demonstrated its potential to teach variations of a task to the robot, and we have used as example the improvement of the selection of grasping points in textiles like polo-shirts.

Taking the experimental results into consideration, vector autoregression provides better results compared to using only FINDDD descriptors. Even though there are results that are unacceptable for human beings, in average, this method moved the grasping point closer to a desired point. Moreover, using the natural hand movements made the robot easier to operate. The Leap Motion sensor gave the user the sensation of real-world grasping without the touching.

The method explained in this work is only tested with spread open shirt without any significant deformations or rotations. However, given enough data to train the system for these cases, the results obtained shows that this method contributes a significant improvement for shirt manipulation. In the future, more data can be collected for different kinds of deformations and the system can be enabled to create better solutions for different deformations on the shirt.

References

[1] A.J. Schmid, N.Gorges, D. Goger and H. Worn. Opening a door with a humanoid robot using multi-sensory tactile feedback. In *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on,* pages 285-291, May 2008.

[2] Siddhartha. Srinivasa, Dave Ferguson, CaseyJ. Helfrich, Dmitry Berenson, Alvaro Collet, Rosen Diankov, Garratt Gallagher, Geoffrey Hollinger, James Kuffner, and MichaelVande Weghe. Herb: a home exploring robotic butler. *Autonomous Robots*, 28(1):5-20,2010.

[3] Christian Bersch, Benjamin Pitzer, and Soren Kammel. Bimanual robotic cloth manipulation for laundry folding. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1413-1419, 2011.

[4] A. Ramisa, G. Alenyà, F. Moreno-Noguer, and C. Torras. Finddd:A fast 3d descriptor to characterize textiles for robot manipulation. In *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on,* pages 824-830, November 2013.

[5] Gabriella Csurka, Christopher R. Dance, Lixin Fan, Jutta Willamowski, and Cedric Bray. Visual categorization with bags of keypoints. In *Workshop on Statistical Learning in Computer Vision, ECCV*, pages 1-22, 2004

[6] Arnau Ramisa, Guillem Alenyà, Francesc Moreno-Noguer, and Carme Torras. Using depth and appearance features for informed robot grasping of highly wrinkled cloths. In *ICRA*, pages 1703-1708. IEEE, 2012

[7] Arnold Zellner, An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association*, 57(298):348-368, 1962

[8] Kosuge, K.; Hirata, Y., Human-Robot Interaction, *Robotics and Biomimetics, 2004. ROBIO 2004. IEEE International Conference on*, pp. 8, 11, 22-26 Aug. 2004

[9] Liying Cheng; Qi Sun; Han Su; Yang Cong; Shuying Zhao, "Design and implementation of human-robot interactive demonstration system based on Kinect" *Control and Decision Conference (CCDC), 2012 24th Chinese*, pp.971,975, 23-25 May 2012

[10] Kun Qian, Jie Niu and Hong Yang , "Developing a Gesture Based Remote Human-Robot Interaction System Using Kinect" *International Journal of Smart Home*, vol. 7, no. 4, July 2013

[11] Bassily, D.; Georgoulas, C.; Guettler, J.; Linner, T.; Bock, T., "Intuitive and Adaptive Robotic Arm Manipulation using the Leap Motion Controller," *ISR/Robotik 2014; 41st International Symposium on Robotics; Proceedings of*, pp.1,7, 2-3 June 2014