Proceedings of the 2017 IEEE
International Conference on Robotics and Biomimetics
December 5-8, 2017, Macau SAR, China

A Study on Human-Robot Collaboration for Table-setting Task

Krati Saxena*, Rollyn Labuguen*, Ravi Prakash Joshi*, Nishanth Koganti†* and Tomohiro Shibata*
*Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology, Kitakyushu, Japan
†Graduate School of Information Science, Nara Institute of Science and Technology, Nara, Japan
Email: {saxena-krati, labuguen-rollyn, joshi-ravi-prakash}@edu.brain.kyutech.ac.jp, nishanth-k@is.naist.jp, tom@brain.kyutech.ac.jp

***Abstract*—As the robot technology is advancing, it is possible to use robots for basic day-to-day chores, so that the burden can be taken off from humans. To make the robots perform such tasks, it is necessary for them to handle different types of objects. Manipulation of deformable objects such as cloth is a challenging task for a robot because of high dimensionality and large number of possible configurations of cloth. Previous studies have covered large number of simple manipulations of cloth articles. In this paper, we are focusing on table setting task that requires putting on a sheet of cloth on the table.**

This paper proposes human-robot collaboration for table-setting task based on visual assessment. We used Baxter robot to hold two corners of rectangular tablecloth and other two corners are held by human. A head-mounted Kinect sensor is used to get the state of cloth and Robot arms are used for controlling the position of cloth corners. We use features from Kinect sensor to assess whether the placement of the cloth is successful or not. We demonstrate an initial study of the system that can achieve promising results towards table setting task through human-robot interaction.

Keywords—Human-Robot interaction; cloth state recognition; cooperative manipulation; service robots

I. Introduction

As the aging population is increasing in the society, automated systems are needed to complete basic tasks. Robots are already being used in various domains to do many chores, with or without the collaboration of humans. For this, robots should be able to manipulate and handle daily life objects. It is observed that handling deformable objects such as cloth and paper can be difficult for the robot as it requires understanding and identifying the objects state, shape and pattern. Very basic cloth manipulation tasks can also be challenging for the robot because of variability in shape, size, colors of the cloth and infinite number of possible configurations.

But many day-to-day tasks do not need a very complex manipulation of deformable objects. One such basic task is putting the tablecloth on a table. Some similar tasks can be placing cloth covers on top of household items such as refrigerator, televisions and even cars. Similarly, bed making comes under the same category of cloth manipulation tasks. These simple tasks can find applications in various places such as hotels, restaurants, hospitals and households. The challenge in the problem mainly lies in efficiency of the task to be performed. Real-time manipulation of the cloth along with the feedback is needed to complete these types of tasks.

Fig. 1. a) Baxter Robot with a head-mounted Kinect sensor is holding the tablecloth, b) The initial setup of tablecloth placement task.

In this paper, we primarily focus on the interaction between human and robot while they are placing a tablecloth such that the cloth covers the table and is wrinkle-free. Cloth state recognition is done using visual information, and position control is used to set the tablecloth in place. Depth information of the cloth has an added value on detecting whether wrinkles are present or not. Experimental results show that this table-setting task is easily accomplished by the collaborative agents.

The outline of the paper is as follows: Section II provides some insights on previous works in this domain. Section III describes the problem statement, system setup and overview of the method used. Section IV, V and VI discusses the methods used in detail. Section VII presents the experiments and results. Finally, we conclude in Section VIII with discussion and possible future aspects.

II. Related Works

Cloth handling and manipulation have attracted many researchers in past few years. Many studies have been done on simple and complex manipulation of clothes [5,6,7,8]. Van den Berg et al. [5] proposed an algorithm for folding the clothes based on their geometry, rather than the physics of the cloth. Miller et al. [6] defined a quasi-static cloth model which allowed to neglect the complex dynamics of the cloth, and use only simple geometry to manipulate and fold the garments.

978-1-5386-3741-8/17/$31.00 © 2017 IEEE

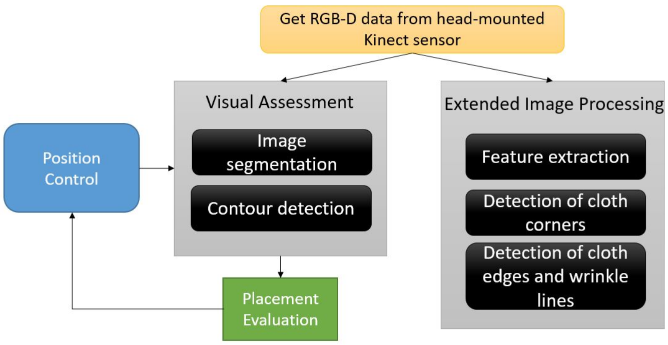

Fig. 2. Workflow of the tablecloth setting task.

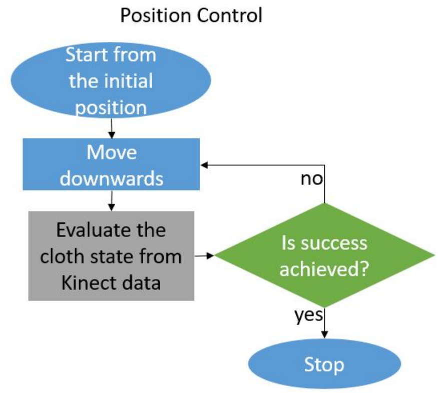

Fig. 3. Position Control of Baxter.

Both studies showed successful implementation on variety of garments. Koganti et al. [7] performed clothing assistance task using Bayesian nonparametric latent space learning. Li et al. [8] proposed an algorithm to find an optimal trajectory to fold the garment, given the start and end configurations.

There have been several studies in the field of human-robot collaboration and manipulation of deformable objects. Kruse et al [1] proposed a method of collaborative manipulation of deformable sheet between a person and a robot, where robot follows the human motion to handle the cloth. They have used RGB-D information to detect folds, joint-torque controller to smooth out the folds and to maintain tension of the sheet. Later, they extended their work [2] by using a mobile dual-arm robot. For this, they determined optimal position and velocity setpoints from human pose information. The robot then attempts to follow these setpoints while maintaining other constraints. Koustoumpardis et al. [3] proposed human robot interaction to fold deformable sheets. Their method was based on force and RGB-D feedback in both higher and lower control levels of the process. In the higher level, the perception of the human's intention was used for deciding the robot's action, then in the lower level the robot reacted to the force/RGB-D feedback to follow the human guidance. Our work attempts to achieve a similar task but here it is specific to placement of the cloth on a table by using visual feedback. The human intention is not required in our task. The robot's actions control the height of the tablecloth and are only dependent on the current state of the cloth, which can be observed through Kinect sensor.

III. Problem Statement and System Setup

We plan to design a system that can efficiently work with deformable objects and complete the basic tasks of placing cloth on any item. The scope of this task is huge but in this paper, we are mainly taking in consideration, the task of placing tablecloth on the table. Initial setup of our experiment is shown in Fig. 1(b).

Since the task is collaborative, it requires two agents, one is the robot and the other is human. We are working with a dual-arm robot, Baxter by Rethink Robotics (as shown in Fig. 1(a)) which is used to perform this task. Initially, robot is grasping two corners and human is grasping the other two corners of the tablecloth in the air. Then, both move the cloth downwards to lay it on the table. The challenge lies in getting the cloth successfully placed, without wrinkles or folds. We use position control to pull the corners of the tablecloth in the downward direction, and Kinect sensor to get the real time image of the cloth. This image is then filtered to detect wrinkles and folds on the cloth.

We use various filters to detect the state of the cloth. Firstly, Region of interest is defined by performing HSV segmentation of the tablecloth. This results in an image that has background removed and only tablecloth is visible in the image. Then a mask is created by thresholding the image, to find the contours of the cloth. This is done to realize whether the cloth is in air or on the table. The contour of the cloth is bigger when it is in air, than when it is on the table, since the table edges will form the new contour now. A bounding box is created to enclose the whole tablecloth. For a simple task of putting the tablecloth, this can be used as success condition, i.e. when the contour size switches from bigger contour of cloth to smaller contour of table edges, the task is done. In our case, we have extended the vision assessment to check for the presence of wrinkles on the tablecloth. For this, we use Gabor filters to extract the texture and edge features. The resulting image is used to detect corners of the cloth and detecting wrinkles.

The corner is detected to estimate the position of arms of robot and human hands, as the corners are clutched by robot arm and human hand. It can be used for failure detection that is discussed later. Also, the lines and edges are detected to find the wrinkles and folds on the cloth. The ultimate aim is to place the cloth successfully on the table by minimizing the number of wrinkles on the cloth.

Position control is done by moving the end-effectors of Baxter downwards. The visual evaluation aids the position controller to continue or stop based on whether the success is achieved or not. A flowchart is depicted in Fig. 3. that describes the working of position control. Robot arms move downwards until the cloth is fully placed on the table.

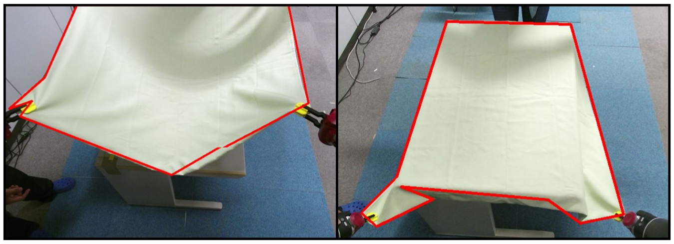

Fig. 4. a) Contours when the cloth is in air, b) Contours when the cloth is placed on the table.

Finally, the depth images of the tablecloth placed on the table are compared to a perfect cloth placement task (without folds and wrinkles) to check how precise the placement task is.

IV. Visual-Assessment

A head-mounted Kinect sensor in Baxter as shown in Fig. 1(a) is used to record data of carrying out the task. Color images are captured and filtered to get necessary data. Various steps involved in this process are discussed below.

A. Image Segmentation

For working with the tablecloth, it is necessary to extract the tablecloth region of the image, as other regions might interfere with the process. In our case, we used clothes of four different colors: lime, olive-green, brick-red and blue-gray. We use HSV segmentation to extract the region of interest. Color isolation can be achieved by extracting a particular hue, saturation, value from the image.

The algorithm consists of two parts: RGB to HSV conversion and applying a threshold mask [11]. In HSV representation of colors, the hue determines the major color, saturation is the intensity and value is the lightness of the image. To isolate the colors, we have to apply masks. A low threshold and high threshold mask for hue, saturation and value is defined. All the pixels lying in the range of these thresholds will be 1, and other will be labelled as 0. For our task, we gave thresholds based on the color of our tablecloth. The values were changed on the basis of required color.

B. Contour Detection

The output of image segmentation is now converted to grayscale and mean adaptive thresholding is applied on it. The image converts to a binary image. Contour detection is applied on this resulting image to find the cloth edges.

Ramer-Douglas-Peucker Algorithm [12] is used for contour detection. It is an algorithm for reducing the number of points in a curve that is approximated by a series of points. It does so by assuming a line between the first and last point in a set of points that form the curve. It checks which point in between is farthest away from this line. If the point (and as follows, all other in-between points) is closer than a given distance 'epsilon', it removes all these in-between points. If on the other hand this 'outlier point' is farther away from our imaginary line than epsilon, the curve is split in two parts: 1) From the first point up to and including the outlier, 2) The outlier and the remaining points. The function is recursively called on both resulting curves, and the two reduced forms of the curve are put back together.

The area and perimeter of the contour is recorded. And a polygon is fitted to show the edges of the tablecloth clearly as shown in Fig. 4. If the tablecloth polygon area and perimeter decreases below a certain threshold, then success is achieved. The vision assessment has been extended for wrinkles and fold detection.

V. Extended Visual Assessment

A. Feature Extraction

For extracting features such as cloth edges, folds and wrinkles, we first applied Gabor filters on the RGB image obtained from the Kinect sensor, as described in [4,9]. Gabor filters are bandpass filters which are used in image processing for feature extraction, texture analysis and stereo disparity estimation because it possesses optimal localization properties in both spatial and frequency domains. The equation is as follows:

$$g(x, y; \lambda, \theta, \psi, \sigma, \gamma) = exp(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2})exp(i(2\pi\frac{x'}{\lambda}\psi)) \tag{1}$$

where,

$$\begin{aligned} x' &= x\cos\theta + y\sin\theta \\ y' &= -x\sin\theta + y\cos\theta \end{aligned} \tag{2}$$

In this equation, λ represents the wavelength of the sinusoidal factor, θ represents the orientation of the normal to the parallel stripes of a Gabor function, ψ is the phase offset, σ is the sigma/standard deviation of the Gaussian envelope and γ is the spatial aspect ratio, and specifies the ellipticity of the support of the Gabor function.

We have considered σ to be constant. Similarly λ, γ and ψ are also constants. θ is varied. Because a Gabor filter has directionality, the resulting images contain various edges that rely on θ settings.

Output after applying Gabor filters is then used for two purposes: (1) detection of corners of the table and the cloth; (2) detection of wrinkles and cloth edges.

B. Detection of cloth corners

We have used Shi-Tomasi corner detector [10] on the resulting image of Gabor filters to detect corners of the cloth. It is determined by a score R, given as,

$$R = \min(\lambda_1, \lambda_2) \tag{3}$$

where, λ_1 and λ_2 are the Eigen values of M. M is defined as follows:

$$M = \sum_{x,y} w(x, y) \begin{bmatrix} I_x I_x & I_x I_y \\ I_x I_y & I_y I_y \end{bmatrix} \tag{4}$$

Fig. 5. Corner detection on tablecloth.

Fig. 6. Failure cases: a) Human held the cloth in air, b) cloth does not cover the table entirely, c) Unever placement of cloth by human.

Here, $w(x, y)$ is the window function. It is either a rectangular window or Gaussian window which gives weights to the pixels underneath. I_x and I_y are image derivatives in x and y directions respectively. If R value is greater than a certain predefined value, it is marked as a corner.

Fig. 5 shows the corner extraction of the tablecloth. In a way, this can be used to track the human hands, since the corners are held by human. This can be used for failure detection, where the failure scenarios are: (1) the human did not move the tablecloth down, (2) the robot pulled the tablecloth too much, as a result of which, the tablecloth from the other side does not cover the table completely. This situation can occur if the size of tablecloth and table top is almost same or the size of cloth is less than that of table as shown in Fig. 6(b). There can be also be cases where human did not put the cloth properly as shown in Fig. 6(c). In the first failure scenario, the corners will not move down or will not be visible if they are out of the camera view range as shown in Fig. 6(a). For other cases, more corners will appear in the areas where the tablecloth does not cover the table.

C. Detection of cloth edges and wrinkle lines

Line Detection is done to find out the edges of the cloth. Basically, this is used to define the success of the task. When the cloth is in the air, the edges of the cloth determines the lines in the filters. When the table cloth is pulled down and is set down on the table, the table edges determines the lines. So this can be formulated as a success scenario for the task. If the lines of cloth switch to the lines of table, the cloth has been placed on it.

For this, the resulting image of Gabor filters is considered. The image is converted to grayscale and mean Adaptive thresholding is applied on the image to change it to binary image. And edges are detected using laplacian edge detection method. Hough Transform is applied to the resulting image to detect the lines.

Fig. 7. a) Wrinkles and edges are seen during placement, b) sometimes no significant wrinkles are observed because human tries to place the cloth properly or light conditions are not appropriate, c) once the cloth is placed properly, the robot stops moving down.

Hough transform is a technique which can be used to isolate features of a particular shape (in this case, lines). It requires the desired features to be specified in some parametric form. The edges points (in the resulting image of edge detection) are mapped to the Hough space and stored in an accumulator. The accumulator is interpreted by thresholding and other constraints to yield lines of infinite length. And then the infinite lines are converted to finite lines, which are then superimposed on the original image. This finally results in the necessary lines in the image. Fig. 7 shows the output image. This also detects wrinkles (smaller lines). Next task of the robot is to minimize the number of wrinkles on the cloth, i.e. minimize the number of visible lines in the image.

D. Calculation of Structural Similarity Index

The depth images of the cloth are used to calculate the index for structural similarity (SSIM). The SSIM index is a method for measuring the similarity between two images. It can be viewed as a quality measure of one of the images being compared, provided the other image is regarded of perfect quality. The SSIM index is calculated on various windows of an image. The measure between two windows x and y of common size $N \times N$ is:

$$SSIM(X, Y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{({\mu_x}^2 + {\mu_y}^2 + c_1)({\sigma_x}^2 + {\sigma_y}^2 + c_2)} \quad (5)$$

where, μ_x and μ_y are the averages of x and y, ${\sigma_x}^2$, ${\sigma_y}^2$ and σ_{xy} are the variances and covariance of x and y. Constants c_1 and c_2 are included to avoid instability when $({\mu_x}^2 + {\mu_y}^2)$ and $({\sigma_x}^2 + {\sigma_y}^2)$ are very close to zero. Specifically, they are defined as $c_1 = (k_1L)^2$ and $c_2 = (k_2L)^2$. Here, L is the dynamic range of the pixel-values, $k_1 << 1$ and $k_2 << 1$. We used $k_1 = 0.01$ and $k_2 = 0.03$, as described in [13]. The SSIM index ranges between -1 and 1, where 1 indicates perfect similarity.

The final depth information in a clothing trial is selected and compared to a successful scenario. SSIM attempts to model the perceived deviation in the structural information of the depth image. This leads to a simple

Fig. 8. a) Successful scenario on Blue-gray cloth, b) random trial with wrinkles on Blue-gray cloth, c) Successful scenario on Olive-green cloth, d) random trial with wrinkles on Olive-green cloth.

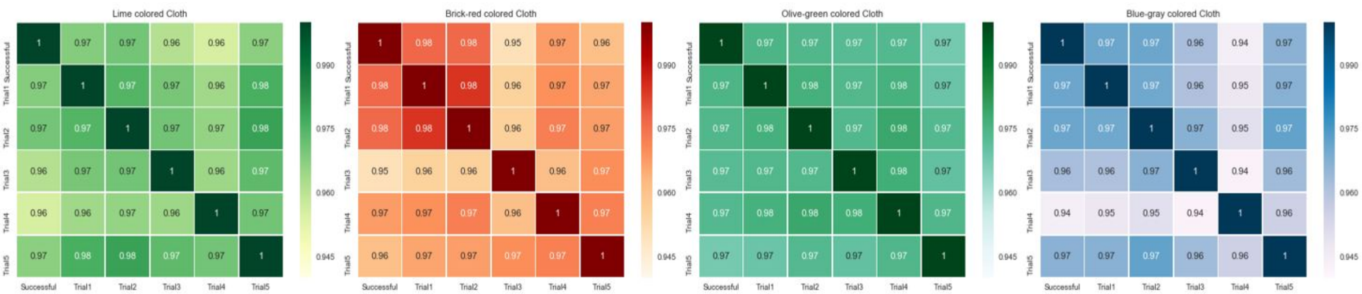

Fig. 9. Heat maps of SSIM index comparision for five trials of cloth placement task to a successful task for a) Lime b) Brick-red c) Olive-green and d) Blue-gray colored tablecloth.

but effective approach to account for flatness and wrinkle formation on the final state.

VI. Postion Control

Baxter moves the cloth down using position control. For this task, we assumed that the Baxter end-effectors are moving only in the z-direction. An initial position is given, where the end-effectors are holding two corners of the cloth in the air, some distance above the table as shown in Fig. 1(b). Baxter pulls the cloth down until it is set on the table.

There are two success conditions here: the first one is from the simple visual assessment. If the contour area and perimeter are decreased below a threshold value at any point of time, it is assumed that the cloth has been successfully placed on the table. The other success condition comes from the wrinkle detection, i.e., if there are no significant wrinkles on the cloth surface, then it is assumed that cloth has been properly set.

Failure scenarios are detected from the corner detection algorithm. If the corners of human side have not moved down, then the cloth is assumed to be in air from the humans side of cloth. So, this constitutes one failure condition. Another failure condition generally occurs if there is no significant difference in the size of the table top and size of the tablecloth. If they are of similar sizes, Baxter may pull the cloth towards itself to remove wrinkles and end up creating uneven placement. The resulting placement will make the cloth hanging on the side of the robot, and the humans side will be partially lying on the table. It is also possible that human did not put the cloth properly. In both cases, the table is not covered completely and position control will stop immediately if multiple corners are detected on humans side of the table.

VII. Experimental Result

The image processing results from the Kinect sensor data is fed to position control to optimally move the robot arms so that success conditions are achieved. Fig. 7 and Fig. 8 shows the extracted features and results of Kinect data. It was observed that the Baxter was able to lay the tablecloth properly on the table each time. In each trial, the system evaluates if the human-robot collaborative task was performed properly.

SSIM index is evaluated on depth images of the cloth. Fig. 9 shows the heat map containing SSIM index comparison for all four types of tablecloths. In the heat maps, ‘Successful’ task is a perfect cloth placement task with no wrinkles or folds. Five random cloth placement trials, ‘Trial 1-5’, are compared to the ‘Successful’ task. SSIM index of the trials depends on the closeness of that trial to the sucessful placement task. The experiment is illustrated in Fig. 8. Fig. 8 (a) and (b) represents the successful task and a random trial of Blue-gray cloth. More wrinkles can be seen in the trial of this cloth. Similarly, Fig. 8 (c) and (d) shows the successful and random trial of Olive-green cloth. Less wrinkles are present in the trial of this cloth.

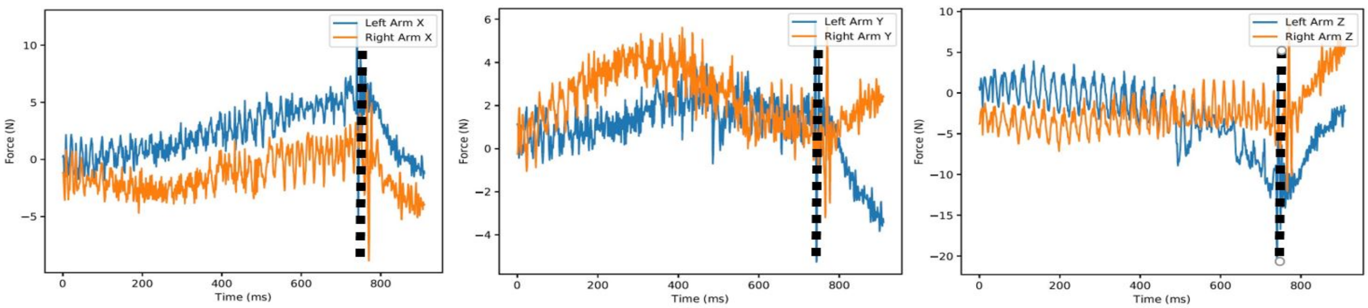

Fig. 10. End-effector forces in x, y and z directions, black dotted lines represents the time when end-effector reaches the table, x-axis represents frames.

The heat maps demonstrates similar results. Since the Olive-green cloth forms less wrinkles and folds, the SSIM index for all trials are relatively closer to 1. On the other hand, Brick-red and Blue-gray cloth forms more wrinkles, so SSIM values for some trials are comapritively lower. Lime cloth has intermediate wrinkle formation.

It is also observed that wrinkle detection has performance constraints if there are improper light conditions. Additionally, the appearance of wrinkles is conditioned on the type and texture of the tablecloth. If the tablecloth is thicker and stiffer, the wrinkles might not appear. Due to human-robot collaboration in this task, we assume that the human always tries to put the cloth perfectly. As a result, there are no significant wrinkle appearances on the cloth, since the cloth is stiff. But folds may appear on the cloth if any agent has not placed the cloth as needed. In the same setting, if the cloth is light and has flexibility like cotton or silk, more winkles may appear on the tablecloth. Patterns on tablecloth may also impact the performance of the Robot. In our setup, we used only monochromatic tablecloths. If there are multi-colored cloth or the cloth has some designs on it, the current system might fail to accommodate the changes. The other factor that can impact the performance is length of the table. If the table is very huge, the human end of the table can get out of range of the robots view. It is possible that the whole table top is not visible in the camera. As a result of this, Baxter will not be able to evaluate the failure conditions as described above.

VIII. Conclusion

In this paper, we implemented an initial step to a human-robot collaborative system that can set a cloth on the table. The robot used position control and visual assessment to complete the task of placing the cloth successfully on the table and at the same time maintain a wrinkle-free surface on the cloth. Fig. 10 shows the forces of the end-effectors of Baxter during the task. It can be observed that as the arms reaches the table, the force starts increasing in all direction. The black dotted line represents the time when the end-effectors reach the table. Each time step represents a frame. Force information can also be used as a condition to detect the success of the task in the future. We plan to extend this work to accommodate the challenges associated with wrinkle detection, light conditions and the type, texture and prints on the tablecloth. This way, there will be no significant impact on the task performance due to above mentioned factors.

Furthermore, our future work is focused on development of sophisticated Robotic Cloth Manipulation system that can understand not only the cloth state and but also the environment. Incorporating the information about the environment will make the system more generalized to perform other similar tasks, e.g. setting the bedsheets and blankets for bed making, covering household items with cloth covers, etc.

References

[1] Kruse, D., Radke, R. J., & Wen, J. T. (2015, May). Collaborative human-robot manipulation of highly deformable materials. In Robotics and Automation (ICRA), 2015 IEEE International Conference on (pp. 3782-3787). IEEE.

[2] Kruse, D., Radke, R. J., & Wen, J. T. (2017, May). Human-robot collaborative handling of highly deformable materials. In American Control Conference (ACC), 2017 (pp. 1511-1516). IEEE.

[3] Koustoumpardis, P. N., Chatzilygeroudis, K. I., Synodinos, A. I., & Aspragathos, N. A. (2016). Human robot collaboration for folding fabrics based on force/rgb-d feedback. In Advances in Robot Design and Intelligent Control (pp. 235-243). Springer, Cham.

[4] Yamazaki, K. (2017). A method of classifying crumpled clothing based on image features derived from clothing fabrics and wrinkles. Autonomous Robots, 41(4), 865-879.

[5] Van Den Berg, J., Miller, S., Goldberg, K., & Abbeel, P. (2010). Gravity-based robotic cloth folding. In Algorithmic Foundations of Robotics IX (pp. 409-424). Springer Berlin Heidelberg.

[6] Miller, S., Van Den Berg, J., Fritz, M., Darrell, T., Goldberg, K., & Abbeel, P. (2012). A geometric approach to robotic laundry folding. The International Journal of Robotics Research, 31(2), 249-267.

[7] Koganti, N., Tamei, T., Ikeda, K., & Shibata, T. (2017). Bayesian Non-parametric Learning of Cloth Models for Real-Time State Estimation. IEEE Transactions on Robotics.

[8] Li, Y., Yue, Y., Xu, D., Grinspun, E., & Allen, P. K. (2015, September). Folding deformable objects using predictive simulation and trajectory optimization. In Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on (pp. 6000-6006). IEEE.

[9] Tutoial on Gabor filters, http://mplab.ucsd.edu/tutorials/gabor.pdf

[10] Shi-Tomasi corner detection algorithm, https://en.wikipedia.org/wiki/Corner_detection

[11] HSV Segmentation algorithm: http://justin-liang.com/tutorials/hsv_color_extraction

[12] Ramer-Douglas-Peucker Algorithm for contour detection: https://en.wikipedia.org/wiki/Ramer-Douglas-Peucker_algorithm

[13] Wang, Zhou, et al. "Image quality assessment: from error visibility to structural similarity." IEEE transactions on image processing 13.4 (2004): 600-612.