Selection of Grasp Points of Cloth Product on a Table Based on Shape Classification Feature

Kimitoshi Yamazaki
Faculty of Engineering
Shinshu University
4-17-1, Wakasato Nagano, Nagano, Japan
kyamazaki@shinshu-u.ac.jp

***Abstract*— This paper describes a method of grasp point detection from an item of cloth with unarranged shape. We focus on the combination of grasp point detector with shape classifier. In the proposed method, Convolutional Neural Network(CNN) is generated for shape classification, and it is also used for extracting a feature vector that presents shape characteristics. Using the feature, grasp points are calculated as image coordinates. Experimental results using real images show the effectiveness of the proposed method.**

***Index Terms*— Grasp points selection, cloth, CNN.**

I. Introduction

Fabric products are indispensable for people to live their lives. Among them, clothes and bedding are used on a daily basis, and we normally maintain and manage cloth products such as washing, drying and storage. There is a need for various work to do. Of these, the washing and drying work has been automated to some extent. Recently the possibility of automation has begun to be seen about storage (folding) work [12]. However, to say the establishment of technology, we still have issues on current situation.

In this paper, we describe grasp points detection from cloth products, which the author considers as one of the problems in automating cloth product operation. For example, assume that there is a cloth product placed in a casual way as shown in the top left panel in Fig. 1. When picking it up and unfolding it, and then moving to folding action, it is necessary to decide which part to grasp.

In the conventional studies related to the automation of the operation of cloth products, there are many approaches to create a situation in which a cloth product is shifted several times so as to be close to a desired shape state. In many cases, the portion to be grasped is easily detected [1] [2]. Kakikura et al. [4] introduced an isolation task using color information. The work was taken to remove the desired cloth product under the premise that each cloth product has a different color, but there was no deep pursuit of the graspping position. Willimon et al. [5] also introduced the task of picking one grasp point for suspending a single cloth product placed on a table casually. Kita et al. [6] proposed a method of matching a deformable shape model for the hanging state with a 3D point cloud measured using a trinocular stereo camera. Osawa et al. [7], Abbeel et al. [8] succeeded in identifying the type of cloth product by a robot observing the contour and the position of the lower end point while operating a cloth product. The advantage of these schemes is that once a robot picks a fabric product up, the robot can deploy the products with a relatively high success rate by successfully selecting some prescribed actions.

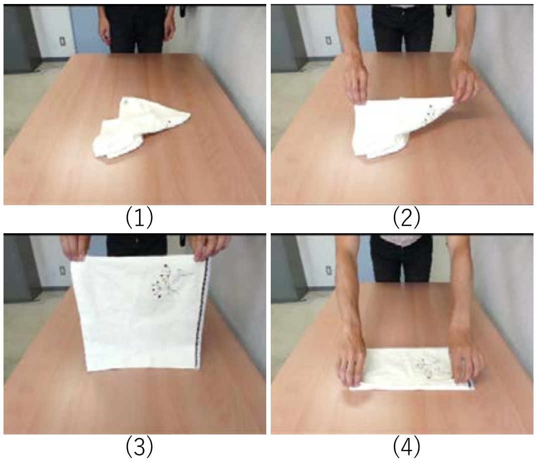

Fig. 1. Picking up an item of cloth by both hands

Meanwhile, if we can unfold cloth products by detecting two or more suitable grasp points from a cloth product in a state that is casually placed on a table, the number of work procedures will be smaller than through a hanging state. It can be said that efficiency is good. However, there is no suitable method as to how to understand the cloth product in a complicated shape state and to perform the grasp point detection. There is an existing research by the author [9], but there are problems such as the success rate of grasped points detection is not sufficient, and it is assumed that hem element can be taken from depth information.

The method proposed in this paper makes it possible to detect an appropriate grasp point from an image. The clue here is the shape state of the cloth product captured in the image. A feature vector expressing the shape state is calculated and the grasp points are obtained by using the calculated feature

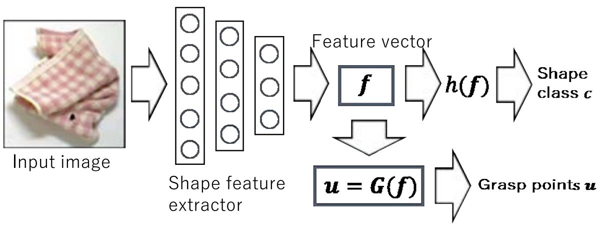

Fig. 2. The outline of the proposed method

vector. As a method of calculating the feature vector, we use convolutional neural network (CNN), classification of shape state and grasp points calculation. At this time, the accuracy of the grasp points detection is improved by taking advantage of the result of the classification. As related work, although a technique of detecting a grasp points using deep learning has been proposed [10], basically it assumed a rigid body. Also, although it can be used for the purpose of finding a part to be gripped independently, like this study it is hard to use for the purpose of finding multiple grasp points spatially separated at the same time.

The structure of this paper is as follows. In Section II, we discuss problem setting and approach. In Section III, the proposed grasp points detection method will be explained. In Section IV, we describe some ideas for improving the performance of the proposed method. Section V describes the experimental results and summarizes it in Section VI.

II. Issues and approach

A. Issues on grasp points selection

As shown in Fig. 1 (1), the cloth product as the target object is assumed to be placed on a horizontal plane such as a table. The goal of this study is to find an appropriate grasp points from here. The grasp points as mentioned here means several places where cloth products can be well unfolded by pinching and lifting at the points, as shown in Fig.1 (2) - (3).

B. Our approach

In order to detect grasp points from a cloth product placed on a desk, it is necessary not only to locate the portion to be gripped locally but also to consider the overall shape state of the cloth product. Detecting only the part to be gripped and its surroundings can not take into consideration the tangling of the cloth away from the grasp point, so it is not always possible to properly unfold it after grasp. Therefore, in this study, we prepare two kinds of parts: one describes the whole shape of cloth products and the other obtains grasp points from them, and cooperate with each other to perform grasp points detection.

Fig. 2 shows the framework of the proposed grasp points detection method. The input is a color image obtained by photographing a cloth product placed on a desk. From the image, one feature vector is calculated. The feature vector is input to each of the part which performs the classification of the shape and the part which calculates the grasp points. The output in this study is the latter result.

III. Grasp points detection based on shape classification feature

In the proposed method, the overall shape and appearance of a fabric product are characterized and used as a clue for detecting the grasp points. This is based on the following assumptions: within a feature vector representing the overall shape, even if the spatial arrangement of the grasp points is apart, a combination of grasp points suitable for unfolding is recorded there. For example, when a corner portion is a grasp point with respect to a rectangular cloth product such as a hand towel, if two corners on the diagonal are grasped, it can not be unfolded in a flat shape. To avoid this, it is necessary to select not only corner detection but also two adjacent corners after understanding the overall shape condition. Also, when a part that is harder to find than a horn is a candidate for a grasp point like a shoulder part of a shirt, expectation is given to a method of estimating a grasp point by using the whole shape as a clue.

The flow of processing will be explained along Fig. 2. When a color image is input, it passes through the feature extractor and a feature vector $\mathbf{f}$ is calculated. In this study, $\mathbf{f}$ is extracted from the layer at the end of the convolutional neural network. Note that this network also serves as a classifier, where $\mathbf{f}$ is input. From the final layer, a numerical value indicating which of the preset classes corresponds to is output. On the other hand, the extracted $\mathbf{f}$ is used for grasp points detection. By inputting $\mathbf{f}$ to the function $G(\cdot)$ obtained beforehand from training data, we obtain a vector $\mathbf{u}$ that arranges the required number of coordinates $\mathbf{u}_i = (u_i, v_i)$ in one column.

Basically, the following linear equation is considered for calculating the grasp point by the function G.

$$\mathbf{A}\mathbf{f} = \mathbf{u} \tag{1}$$

Hereinafter, the matrix $\mathbf{A}$ is called a transformation matrix. The transformation matrix is obtained beforehand from training data. The calculation method is as follows. First, various images of cloth products are collected. For each image, a subject selects and records the grasping points considered to be appropriate.

Then, using the convolutional neural network, a feature vector $\mathbf{f}_k$ for each image is calculated. Next, a transformation matrix is calculated from two matrices: a matrix $\mathbf{F}$ that feature vectors $\mathbf{f}$ are vertically arranged and transposed, and a matrix $\mathbf{U}$ that corresponding grasp point coordinates $\mathbf{u}_k$ are arranged vertically. That is,

$$\mathbf{A}^T = \mathbf{F}^{+}\mathbf{U}, \tag{2}$$

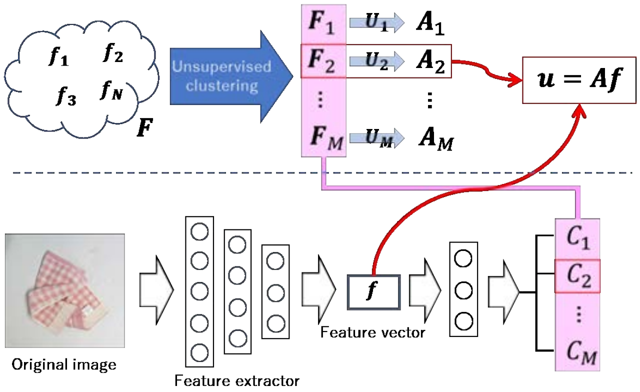

Fig. 3. The improved outline using classification result

where T indicates transpose operation. $\mathbf{F}^+$ is a pseudo inverse matrix of $\mathbf{F}$.

IV. Improvement of Method

A. Adjustment of shape classifier

In the above description, it is assumed that there is a trained convolutional neural network. This network might consist of only a dataset used in this study, but it might also use the one for image classification learned by a large amount of training data already. Since it requires a large amount of training data to learn a multilayered network, the latter method seems to be realistic. Therefore, the latter is assumed in this study.

However, there is room for further thought about the output of the image classifier. We focus on the effectiveness of considering the overall shape of fabric products, so here we consider a classifier that classifies according to the shape. That is, given a learned convolutional neural network, a class defined based on the shape of the fabric product is given and metastatic learning is performed. Based on the network obtained as a result, class classification and calculation of feature vectors are performed.

Various methods are conceivable about the definition of the shape class of the cloth product. A simple way is to apply unsupervised clustering to training images of cloth products and assign the same labels to images that belong to the same group as criteria for classification. If the data to be input to this unsupervised clustering is the feature vector $\mathbf{f}$ obtained from the terminal layer of the network, the above procedure can easily be executed. Furthermore, if we want to consider grasp points $\mathbf{u}$ in clustering, we create a new vector that concatenates $\mathbf{f}$ and $\mathbf{u}$ and uses it as input to unsupervised clustering.

B. Selection of transformation matrix by shape classification result

In Section III, we explained the method of calculating one transformation matrix from all the training data. However, since the transformation matrix determined in this way has a large size and contains various shape information, the accuracy of the grasp points detection can not be expected in some cases. Therefore, in conjunction with the classification result of the shape class, we reduce the amount of calculation and improve the accuracy.

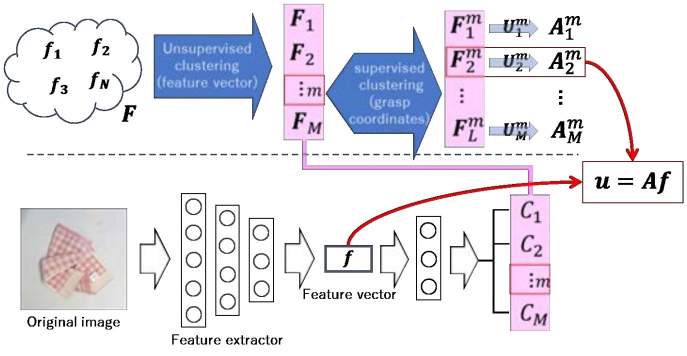

Fig. 4. The improved outline using two types of classification result

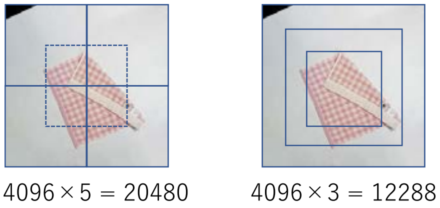

Fig. 5. Two types of extentions for feature description

Figs. 3 and 4 show the outline. The former is a case where clustering is performed from only feature vectors. The latter case is a case where clustering is performed from vectors based on a combination of feature vectors and grasp point coordinates. As we will discuss in the next section, the evaluation by our dataset showed that the latter performance was somewhat better.

When constructing a shape classifier, set the number of clusters as N and divide the training data by unsupervised clustering. Instead of calculating one transformation matrix from all the training data, using only feature vectors included in each cluster $c_i (i = 1, 2, \ldots, N)$, calculate $\mathbf{A}_i$.

Then, when input data is given, the result of shape classification is acquired. If its output is c_i, grasp point detection by Eq. (2) is done using $\mathbf{A}_i$ (Fig. 3 shows the case of $i = 2$. In this method, if the shape classification fails, the estimation accuracy might decrease, but if the classification is successful, the position accuracy of the grasp points detection can be greatly enhanced.

C. Extension of feature expression

We have assumed that one feature vector is calculated from one image. In this subsection, we extend the calculation method and describe a method for enhancing the expressive power of feature quantities.

The first method is to divide the input image into rectangles and calculate feature vectors from each rectangular region. In this study, as shown on the left side of Fig. 5, feature vectors were calculated with a total of five rectangles, rectangle divided into 4 and rectangle placed in the center.

The second is the calculation of feature vectors on multiple scales. As shown on the right side of Fig. 5, a rectangular area having a different size from the center of the input image is set, and a feature vector is calculated for each rectangle. In this study, calculations were made on three scales.

In both methods, a plurality of calculated vectors are connected and used as one feature vector after that.

V. Experiments

A. Experimental settings

A horizontal table top board was prepared, and a camera was installed about 700 [mm] just above the board. For the camera, a camera module FCB-M made by SONY Inc. was used. The size of color image was 640×480 [pixel]. Under this experimental condition, one pixel was equal to approximately 1 millimeter, so subsequent results are pixel notation but can be considered as millimeter.

We adopted AlexNet [11] as a composition of convolution neural network. As the initial value of the weight, the learning result by ILSVRC2012 image data set [13] was used. Then fine-tuning of end-to-end was carried out using data set of shape class of cloth product. Feature vectors were extracted from the obtained network and grasp points were calculated by the method described in Sections 3 and 4. The number of elements of the feature vector $\mathbf{f}$ obtained from the feature extraction unit was basically 4096. When doing the extension described in IV-C, as shown in Fig. 5, it was 20480 dimension and 12288 dimension, respectively. Two grip points were selected from one image: That is, the number of elements of $\mathbf{u}$ was four. When learning the network, k-means method was applied to the generation of the identification class.

In evaluating the experimental results, the definition of the error in this experiment was defined as shown in Fig. 6. If the error value is near to $(0, 0)$, it means that desired grasp points are obtained.

B. Experiments using a cloth product

A rectangular hand towel was placed on a table casually, a dozen of images was then taken, and two grasp points per one image were instructed and recorded. Some images are shown in Fig. 7. Each captured image was rotated / translated to increase the data. Two types of data sets were created. One dataset, named data set A, was translated every 10 pixels in the range of 70 pixels, and was rotated every 5 degrees in the range of 50 degrees. The other data set, named data set B, was translated every 20 pixels in the range of 80 pixels and was rotated every 20 degrees in the range of 360 degrees. The number of increased data was about 5000. From each dataset, 100 images were extracted for testing, and a classifier and a grasp points detector were trained from the remaining data.

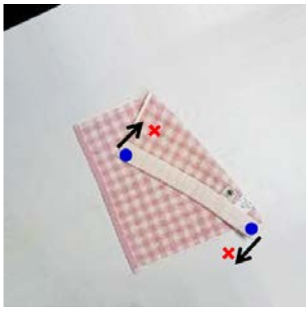

Fig. 6. Definition of error. Difference vectors from the taught grasp points are regarded as errors.

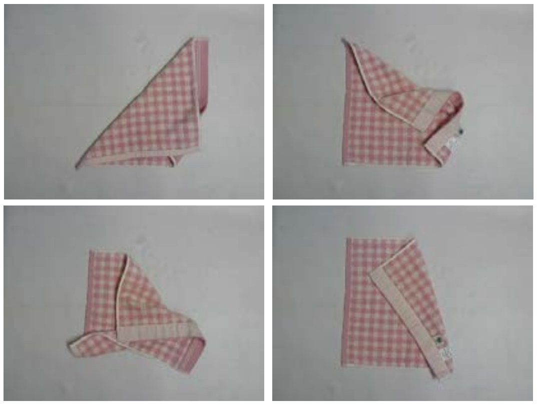

Fig. 7. Handtowel Images

For dataset A, when using the multi-scale type feature vector with the number of clusters as 50, it was possible to detect the grasp point with an error of approximately 5 pixels or less. For dataset B, Fig. 8 is a plot of the error. Here, the blue dots are the result of not using the classification by shape (shown in Fig. 2), and the green points is the result of the detection of the grasp points by the configuration of Fig. 3, with each of the 50 clusters as the classification criteria. From this result, it can be seen that the performance of the grasp points detection is greatly improved when the classification result is used.

Fig. 9 is the result of using three types of feature vector (one is original, and others are explained in Section IV-C). The red dot was the result when using the feature vector of

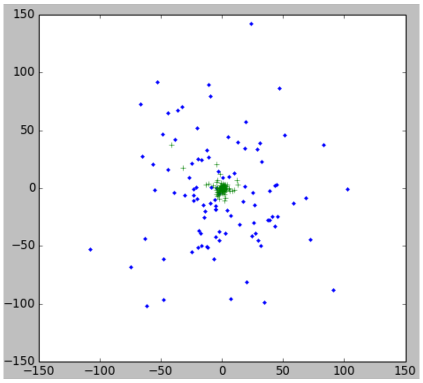

Fig. 8. A Comparison of calculation methods. Blue: Using a single $\mathbf{A}$ for grasp point calculation, Green: Using $\mathbf{A}_i$ depending on the result of shape classification.

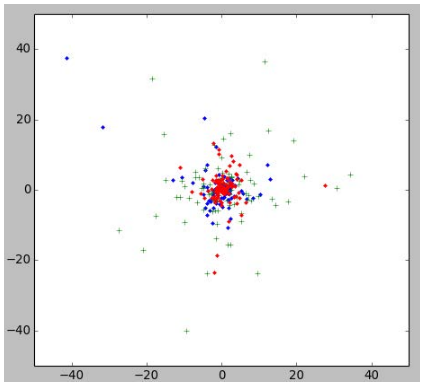

Fig. 9. Grasp point error for handtowel images by three types of feature representation.

multi-scale representation, and the variation was the smallest.

C. Experiments using different fabric products

Experiments on grasp points selection were performed on the four kinds of cloth products shown in the upper four figures of Fig. 10. Place the fabric product in a folded state and select two grasp points as indicated by the red dot in the figures. In addition, as shown in the lower row, other images were captured while gradually shifting the folded portion. Furthermore, as in the previous section, new images was generated by rotating and translating the obtained images, and data was increased. Concretely, parallel translation was performed by 20 pixels or 80 pixel in the longitudinal direction and lateral direction, and after each parallel movement, a rotation of 20 degrees was added. For instructed grasp points, the same rotation and parallel movement were added and recorded. The final number of data (an image and a pair of grasp points) was about 9000. From this dataset, 100 images were extracted for testing, and a classifier and a grasp points detector were trained from the remaining data.

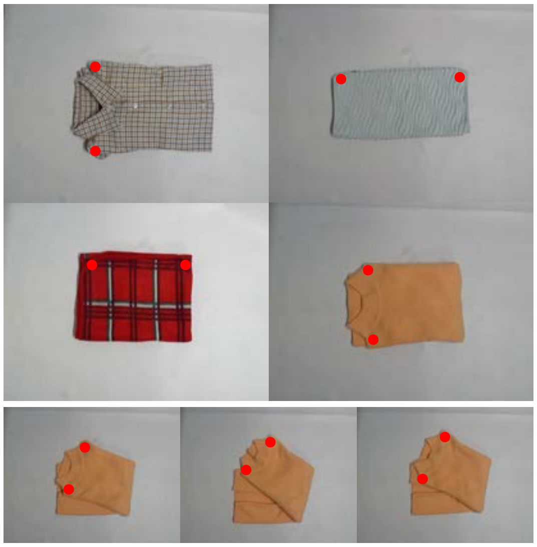

Fig. 10. Original images used for evaluation

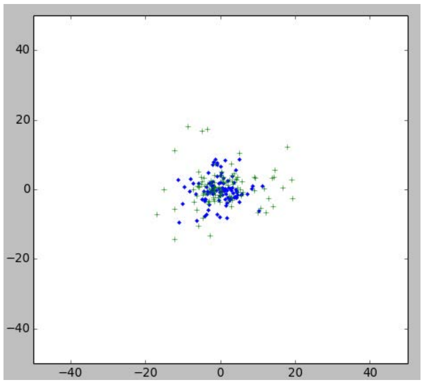

Fig. 11. A result of grasp point error for 4 types of clothes

Fig. 11 shows the result for the 100 test images. The experimental conditions were as follows. The number of clusters was 50, and feature vectors were calculated by multi-scale manner (shown in the right panel in Fig. 5). The blue dots in Fig. 11 show a result when the translation amount at the data padding was set to 20 pixels. On the other hand, the green points in the figure is the case where the translation amount was expanded to 80 pixels.

In both cases, the error was approximately 10 pixels or less, and good results were obtained. This means that the grasp points could be calculated with less than half of the deviation that was originally given.

D. Discussion

In the experiments described above, there were not many types and shapes of cloth products, but in that range, the grasp points could be detected with generally usable accuracy. In Section V-C, cloth products of different patterns, types, fabrics were collected and used as one training data, but grasp points could be obtained without being affected by differences in cloth products. On the other hand, in Section V-B, we found that the proposed method has applicability to fabric products of various shapes. In these results, the error of the grasp points detection becomes 10 pixels or less, and this value is half of or lower than the parallel movement interval at the time of increasing the data, so it is considered that a certain performance was obtained.

However, when shapes significantly different from any of the shapes of the training data are input, it can not deal with the situation at present. As a countermeasure, it is one way to increase the training data so that a shape close to any shape can be found in the training data. On the other hand, if we can define a feature expression that interpolates an unknown shape between two known shapes, we can reduce the amount of training data.

In addition, in the experiment described above, the feature vector of 4096 elements was extracted from CNN, but this vector does not necessarily have sufficient expressive power. In addition to this, two kinds of expansion were performed, but the effect was limited. These features are obtained from the upper layer of the network, and there is a possibility that the overall shape is abstracted too much from the standpoint of grasp points detection. As a countermeasure, we also considered a method of extracting elements from all layers. It will be possible to obtain expressions combining direct information such as local edges and abstract information such as overall shape, which might improve the accuracy of grasp point detection.

In these experiments, rotation and parallel translation were added to the original images. As a result, there were many cases where the given grasp points goes out of the image region. However, even in such a case, the grasp points could be detected with the above-mentioned accuracy. That is, even when the shape of the cloth product is partially missing, there is a possibility that an appropriate grasp points can be determined by calculating a feature vector from a visible shape and utilizing it.

VI. Conclusions

In this paper, we described a method of detecting appropriate multiple grasp points by using the overall shape of a cloth product. The proposed method consists of a shape classifier and a grasp points calculator, and performs shape classification and grasp points detection using a feature vector expressing the overall shape. Regarding grasp points detection, we introduced some ideas that contribute to performance improvement, such as by using the result of shape classification, and confirmed its effect by experiments using actual images. In the scope of these experiments, it was possible to perform grasp points detection with an error of about 10 [pixel].

In this paper we assumed to use convolutional neural network, but the framework of the proposed method is not limited to it. If there is a feature extractor that can express the overall shape of the fabric product, it is easy to replace it.

For future work, we evaluate by applying the proposed method to more fabric products. In addition, it is also necessary to actually perform unfolding behaviors by robots using the result of the grasp points detection.

Acknowledgement

This work was partly supported by NEDO and JSPS KAKENHI Grant Numbers 26700024.

References

[1] A. Doumanoglou, A. Kargakos, T. Kim, S. Malassiotis: "Autonomous Active Recognition and Unfolding of Clothes using Random Decision Forests and Probabilistic Planning," in Proc. of Int'l Conf. on Robotics and Automation, pp. 987 - 993, 2014.

[2] Hiroyuki Yuba, Solvi Arnold and Kimitoshi Yamazaki: "Unfolding of a Rectangular Cloth Based on Action Selection Depending on Recognition Uncertainty," in Proc. of IEEE/SICE Int'l Symp. on System Integration, pp. 623 - 628, 2015.

[3] S. Cuén-Rochín, J. Andrade-Cetto and C. Torras: "Action Selection for Robotic Manipulation of Deformable Planar Objects," in Proc. of Frontier Science Conference Series for Young Researchers: Experimental Cognitive Robotics, pp. 1 – 6, 2008.

[4] K. Hamajima and M. Kakikura: "Planning Strategy for Unfolding Task of Clothes – Isolation of clothes from washed mass –," in Proc. of Int'l. Conf. on Robots and Systems, pp. 1237 – 1242, 2000.

[5] B. Willimon, S. Birchfleld, I. Walker: "Model for Unfolding Laundry using Interactive Perception," in Proc. of IEEE Int'l Conf. on Intelligent Robots and Systems pp. 4871 - 4876, 2011.

[6] Y. Kita, F. Saito and N. Kita: "A deformable model driven method for handling clothes," Proc. of Int. Conf. on Pattern Recognition, Vol.4, pp. 3889 – 3895, 2004.

[7] 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.

[8] J. Maitin-Sphepard et al.: "Cloth Grasp Point Detection based on Multiple-View Geometric Cues with Application to Robotic Towel Folding," Int'l. Conf. on Robotics and Automation, pp.2308 – 2315, 2010.

[9] Kimitoshi Yamazaki: "Grasping Point Selection on an Item of Crumpled Clothing Based on Relational Shape Description," in Proc. of IEEE/RSJ Int'l Conf. on Intelligent Robots and Systems, ISBN: 978-1-4799-6934-0, 2014.

[10] Ian Lenz, Honglak Lee, Ashutosh Saxena: "Deep Learning for Detecting Robotic Grasps," International Journal of Robotics Research (IJRR), vol. 34, no. 4-5, pp. 705-724, 2015.

[11] Alex Krizhevsky and Sutskever, Ilya and Geoffrey E. Hinton: "ImageNet Classification with Deep Convolutional Neural Networks," Advances in Neural Information Processing Systems 25, pp. 1097–1105, 2012.

[12] Panasonic Laundroid: https://laundroid.sevendreamers.com/ (viewed 2016/11/1)

[13] ILSVRC2012 image dataset: http://www.image-net.org/challenges/LSVRC/2012/(viewed 2017/5/1)