Category Classification of Deformable Object using Hybrid Dynamic Model for Robotic Grasping

Yew Cheong Hou
Institute of Informatics and Computing in Energy
Universiti Tenaga Nasional
Kajang, Selangor, Malaysia
e-mail: ychou@uniten.edu.my

Khairul Salleh Mohamed Sahari
Institute of Informatics and Computing in Energy
Universiti Tenaga Nasional
Kajang, Selangor, Malaysia
e-mail: khairuls@uniten.edu.my

Dickson Neoh Tze How
Department of Electrical and Electronics
College of Engineering
Universiti Tenaga Nasional
Kajang, Selangor, Malaysia
e-mail: dickson@uniten.edu.my

Leong Yeng Weng
Institute of Informatics and Computing in Energy
Universiti Tenaga Nasional
Kajang, Selangor, Malaysia
e-mail: ywleong@uniten.edu.my

***Abstract*—This work studies the problem of classification of a hung garment in the unfolding procedure by a home service robot. The sheer number of unpredictable configurations that the deformable object can end up in makes the visual identification of the object shape and size difficult. In this paper, we propose a hybrid dynamic model to recognize the pose of hung garment using a single manipulator. A dataset of hung garment is generated by capturing the depth images of real garments at the robotic platform (real images) and also the images of garment mesh model from offline simulation (synthetic images) respectively. Deep convolutional neural network is implemented to classify the category and estimate the pose of garment. Experiment results show that the proposed method performs well and is applicable to different garments in robotic manipulation.**

Keywords-garment recognition; modeling and simulation; convolutional neural network; real and synthetic images

I. INTRODUCTION

Service robots are gaining attention in robotic research for housework and elderly care when the robots are increasingly improved to be more intelligent and autonomy. The service robot should provide the ability to assist human in the daily activities in terms of performing repetitive tasks with consistent, precise and reliable. For home service robot, they should be different compared with an industrial robot which needs more functionalities and special abilities to handle different types of house chores. However, to teach a home service robot to handle these deformable objects is a challenging problem due to the variety of sizes, textures, and poses for these objects. In doing housework using a service robot, bringing a piece of crumpled garment into the desired configuration requires the understanding of the garment perception and also the manipulating path planning strategy.

In the dexterous robotic manipulation, classification and pose recognition of the deformable object is considered as a primary step to identify the configuration of the object before any robotic manipulating action to be taken. The computer vision techniques integrated with the machine learning algorithm through garment perception is required that capable of the service robot to handle these deformable objects dexterously by recognizing their deformation. In this paper, we are focusing the research of classify the category and recognizing the pose of hung garment grasped by single robotic manipulator. The overview for the dexterous garment handling in robotic manipulation can be depicted as shown in Figure 1. The main idea of this work is to generate a dataset of hung garment that combines the real and synthetic images from the real garment and deformable model respectively. This is the extended work from our previous researches discussed in particle-based polygonal model [1] and self-generated dataset of hung garment [2]. By applying the particle-based polygonal model into the real garment that crudely spread-out on the flat platform, different type and size of garment mesh model can be constructed and simulated by using our dynamic programming algorithm. The dataset will take advantages of this algorithm to generate large number of data for learning purposes which able to reduce the time cost for design a new garment model in the software. The main contributions in these works are:

- The simple 2D garment mesh model is extracted from the real garment that crudely spread-out on the flat platform. The multiple poses of hung garment mesh model by gravity are simulated and different viewpoints are recorded.
- The garment mesh model is not only used to facilitate folding task but also can be used to estimate the category and pose of the garment. The dataset is generated from the synthetic data of hung garment mesh model integrated with the real data, which later trained in CNN to classify the garment's category and estimate its pose in arbitrary positions.

978-1-7281-4620-1/19/$31.00 ©2019 IEEE
DOI 10.1109/CRC.2019.00021

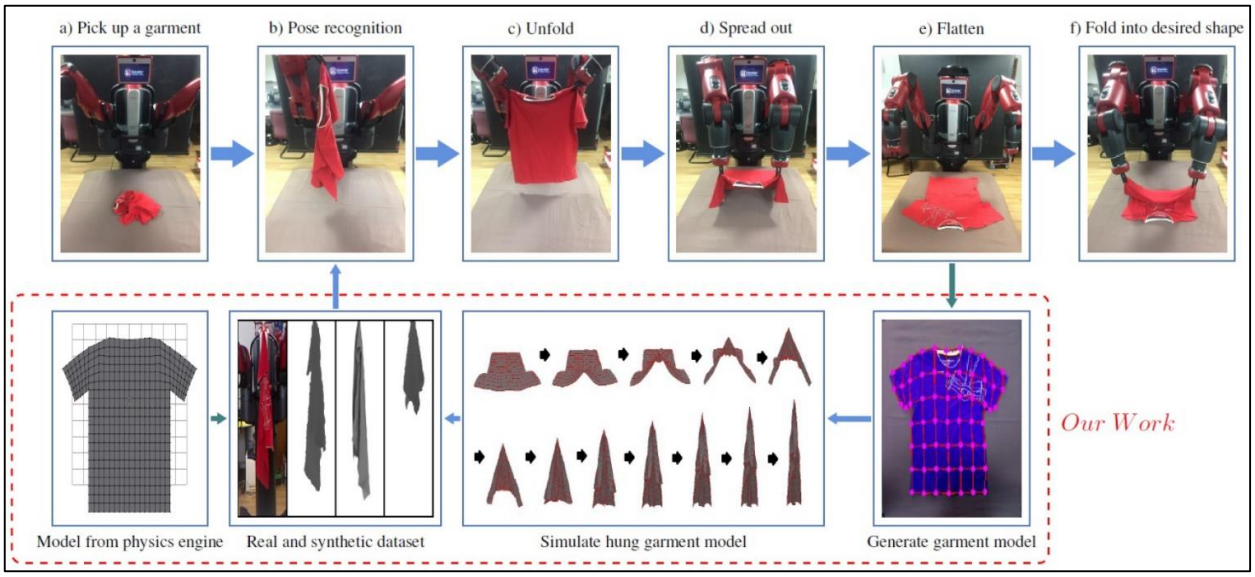


Figure 1. The pipeline for dexterous garment folding in robotic manipulation.

II. Related Work

There has been a significant amount of previous research work on the detecting and manipulating of deformable objects. In the manipulation of deformable objects studied in the literature, Osawa et al. proposed an approach to unfold a garment using two robotic arms [3]. The lowest point of hung garment is selected as priority robotic grasping point and the process is repeated until the reference shape of garment is defined to determine the second appropriate point to unfold the garment on the flat platform. Salleh et al. proposed an approach to identify the second corner of cloth by tracing from first corner of cloth until the end of edge of cloth [4].

When the robot had seen significant advancement in recent years and gaining autonomy and intelligent, the research of service robots getting attraction and most of them focused on housework, restaurant, grocery shop, and maintenance. Maitin-Shepard et al. proposed to identify the corners of a piece of towel using geometric cues [5]. The sequence of grasping the lowest point of towel is suggested and the corner of towel can be identified after several grasps. Miller et al. proposed an approach to handling deformable object by using the polygonal model to understand garment perception [6]. The garment's contour is matched by a best fitted parameterized polygonal model then the category and its configuration can be determined based on the fitted model. Cusumano et al. proposed an approach to unfold different types of garment [7]. In their work, the lowest point of the hung garment is re-grasped repeatedly until the garment's configuration can be determined and modeled by Hidden Markov model. Triantafyllou et al. proposed an approach to unfold a piece of garment using dual-arm manipulator by identifying the two key points on the garment's outline [8]. By assuming the garment is placed on the flat surface, shape matching analysis techniques are applied to match the real garment with a set of foldable templates. The two robotic grasping points that are corresponded to the template can be determined to unfold the garment.

There are several research works are focused on the classification and pose recognition of garment. Kita et al. proposed a dataset of deformable models by constructing the 3D observed garment model to recognize the pose of garment [9]. The silhouette matching technique is applied to obtain the best-fitted garment model from the dataset. Some researchers later proposed different kind of artificial intelligence algorithm to train the real and synthetic dataset in garment recognition by using a Kinect depth sensor or stereo cameras.

Doumanoglou et al. proposed data-driven methods for garment recognition using Random Decision Forests. [10] The depth images of hung garment are collected and the two grasping key-points are estimated by using Hough forests. By using a dual arms industrial robot, they proposed a complete pipeline to handle different categories of garment from their research works [11]. Mariolis et al. used Convolutional neural networks (CNNs) to recognize the pose of garment. They achieve a recognition rate of 89.38% in their recognition of towels, shirts, and trousers [12].

Li et al. proposed a pipeline of laundry works in garment handling process [13]. Firstly, they proposed a model-driven approach using SIFT description for the synthetic dataset of a hung garment in garment recognition. The hierarchy layers of Supoort Vector Machine (SVM) is proposed to recognize the pose of hung garment [14]. Later, their recognition method is applied to the dataset of hung garment using Kinect Fusion technique [15]. The 3D garment model is constructed by rotating the garment 360 degrees vertically and the volumetric features are compared and matched with their offline dataset which the powerful computational resources is required. Similar to previous work, Corona et al. proposed an approach of using a hierarchy of CNNs to train the real and synthetic to classify the category of garment, first grasping point and second grasping point. The promising results are reported in

garment recognition, most of the errors are due to the unsuccessful robotic grasping.

Similar research works to ours is discussed in Mariolis et al. [12] and Corona et al. [16]. We are using depth images of the hung garments as input images for deep convolutional neural networks. Adopting the hierarchical approach used in previous work, our CNNs classifier contains two layers. In the first layer, the CNN model is used for classifying the hung garment to one of the examined categories, which in our case are towels, shirts, and trousers. In the second layer, another deep CNN with similar architecture is employed for inferring the pose of the hung garment by estimating the grasping point location on a garment template. In the second layer, different CNNs are used for inferring the pose of garments belonging to different categories. Hence, the output of the first layer is used for selecting the appropriate CNN for the second layer.

III. METHODOLOGY

The objective in this work is to develop the garment perception enabling a home service robot to execute different types of housework such as garment folding. To fold a piece of garment into the desired configuration, the category and pose of the garment should be identified as the first step either in the condition of picking up under gravity or laying on the flat surface. When a piece of garment is picked up arbitrarily, there are many possibilities may appear by its configuration from visual observation. A large number of exemplars can be deformed for a piece of garment when arbitrarily grasped and hung under gravity as shown in Figure 1(b). Recording all appearance for the deformation of garment is very time consuming and large space storage are required. In deep learning, one of the common problems need to be solved is to collect large-scale data into the desired format. Collecting the data for different categories of a real garment is very time-consuming particularly encompassing the different types, sizes and material properties with its deformation. Hence, a training set that gathers relevant information of garment which can be used in visual recognition purposes is required. The features such as colors and textures of the garments are less significant in this work. To construct the dataset of hung garment, there are three alternatives will be discussed in this section and can be summarized as shown in Figure 2 [2].

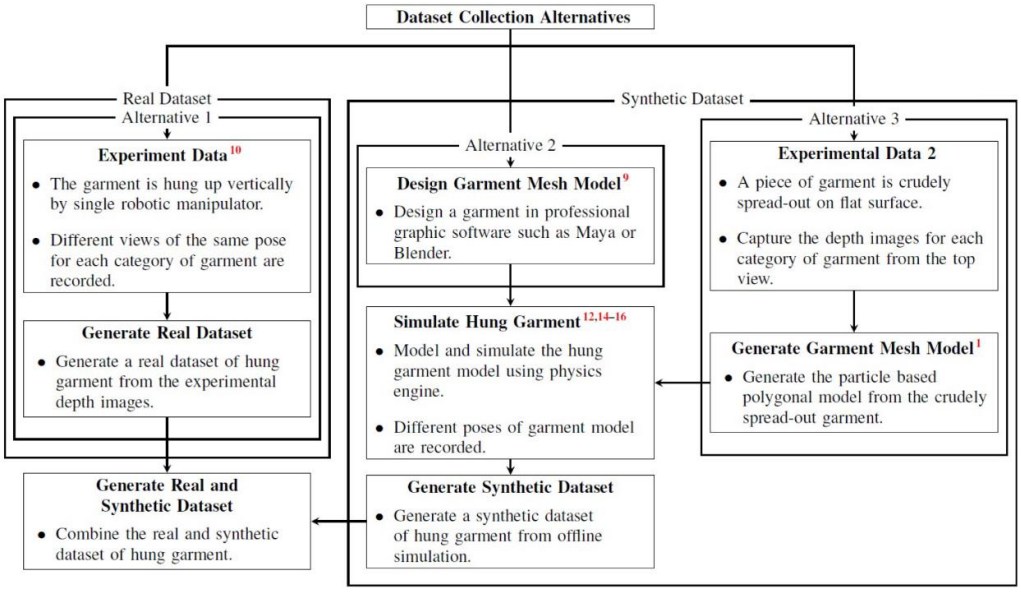


Figure 2. The alternatives for hung garment dataset collection in robotic manipulation [2].

A. Depth and Synthetic Images Acquisition

In this experimental setup, two Kinect depth sensors and one Baxter collaborative robot are used in order to acquire the color and depth input images as shown in Figure 3. One of the depth Kinect sensors is mounted on the head of the robot in order to capture the images from the top view. Meanwhile, the second depth Kinect sensor is mounted horizontal and in front of the robot, the hanging garment is grasped and positioned at the distance approximately 1.2m from the depth sensor. For each real garment, a set of robotic grasping points are predefined as shown in Figure 4. For Alternative 1, a single robotic manipulator starts grasping predefined points and rotating slowly around the vertical axis. A series of depth images are captured for each point, the robot rotates the garment vertically with 360 degrees and the Kinect sensor captures the image of the rotated garment by increasing one rotating degree until full rotation is achieved. Around 360 depth images for each grasping point of real garment can be acquired.

Besides from the real dataset obtained from the depth images, the synthetic dataset can be constructed from a simulation in virtual environment either using professional graphics software or dynamic programming algorithm. For Alternative 2, the professional graphical software such as Maya and Blender can be used to design a new garment model which can encompass different categories and sizes. A set of

points in the garment mesh model can be predefined, the location for each point in the garment model is set as similar as possible to the real garment. For each grasping point, it is selected as constraint and the deformation of the garment model hanging by single robotic manipulator under gravity can be simulated in virtual environment. By applying different physical properties of the garment mesh model, the model look visually similar to depth images can be produced from the virtual simulation. An amount of the depth images that simulate the deformation of hung garment can be acquired. A series of images are captured for each pose of garment, the garment model is rotated vertically with 360 degrees and the virtual camera captures the image of the rotated garment by increasing one rotating degree until full rotation is achieved as shown in Figure 5.

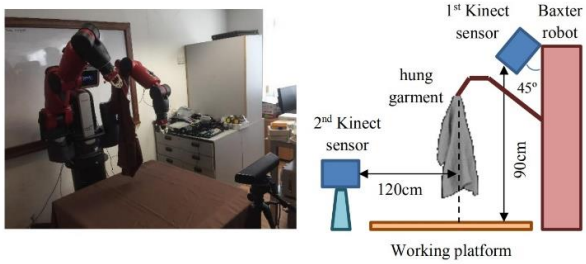


Figure 3. The first Kinect sensor is mounted on the head of Baxter robot. The second Kinect sensor is located in front of the Baxter robot to recognize the pose of garment.

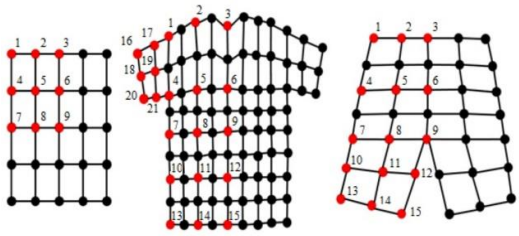


Figure 4. Selected particles defining the poses of the hung garments. Red markers define the poses on the real garment from the particle-based model.

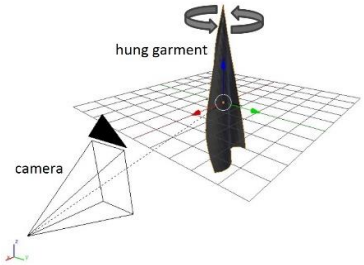


Figure 5. Virtual camera system captured the image of garment mesh model horizontally.

For Alternative 3, the particle-based polygonal model is used to extract the 2D garment mesh model from a piece of crudely spread-out garment [1]. For the garment categories such as shirt and trousers, they are able to discretize into several parts as shown in Figure 6. After the rectangle grid are fitted for its contour, a completed garment mesh model that similar to the contour of garment can be generated. The advantages of this approach is feasible and time efficient once the category of garment was identified. Different types and sizes of 2D garment mesh model can be generated from the crudely spread-out garment. The deformation poses of the garment model hanging by robotic manipulator can be simulated as similar to Alternative 2. Thereafter the simulated hung garment models can be recorded into the synthetic dataset.

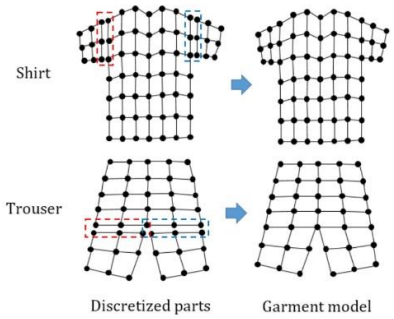


Figure 6. Model combination between discretized parts of garment to form a completed particle based polygonal model.

B. Deep Learning Model for Classification and Pose Recognition

For classification and pose estimation, the Kinect sensor that placed horizontal and in front of the robot is used to capture the depth images of the hung garment as the input images in CNNs. The input to the CNNs is 340×200 pixels depth image. The network structure for recognition in CNNs in this work is depicted in Figure 7. There are five layers in this network structure, the first three layers are the convolutional with pooling and followed by two fully-connected layer. The last layer is connected to the Softmax layer to estimates the output predictions. As shown in Figure 7, feature maps from previous layers are convolved with learnable kernels. The output of the kernels go through either a linear or nonlinear activation function to form the output feature maps. Each of the output feature maps can be combined with more than one input feature map. In general, the convolution can be described as:

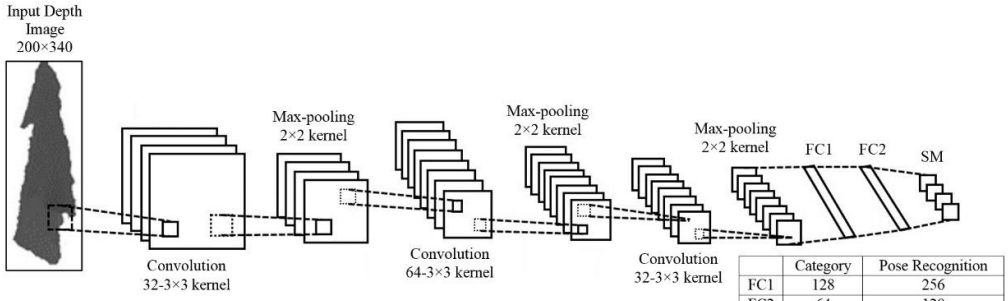
$$x_j^l = f\left(\sum_{i \in M_j} x_i^{l-i} * k_{ij}^l + b_j^l\right) \quad (1)$$

where x_j^l is the output of the current layer, x_i^{l-i} is the previous layer output, k_{ij}^l is the kernel for the present layer, and b_j^l is the biases for the current layer. M_j represents a selection of input maps. For each output map, an additive bias b_j^l is given. However, the input maps will be convolved with distinct kernels to generate the corresponding output maps. The output maps finally go through a linear or non-linear activation function. The fully connected layers (FC) are applied to measure the score of each label from the extracted

features in the convolutional layer from the preceding steps. The final layer feature maps are represented as vectors with scalar values then passed to the fully connected layers. The fully connected feed-forward neural layers are used as a Softmax classification layer (SM). The final layer of the model is a Softmax layer to compute the final prediction in classification. The Softmax function can be defined as:

$$softmax(x)_i = \frac{\exp(x_i)}{\sum_{j=1}^{n} \exp(x_j)} \quad (2)$$

In this work, there are totally four different CNN models with similar structure are used for category and pose recognition of hung garment as shown in Figure 8. For category classification, the Softmax for output prediction is 3 in this network since only three garment categories are tested in this work. The network structure of pose recognition is similar to the network structure in the category classification. The pose estimation is performed by means of classification, with each pose defined as a grasping point for each garment's model. There are 21 poses for shirts, 15 poses for trousers and 9 poses for towels in this work. These poses have been defined to coincide with the particles of the simulated models after discarding symmetric counterparts as shown in Figure 4. The Softmax of the pose estimation is represented by the number of poses and depends on the garment category.



	Category	Pose Recognition
FC1	128	256
FC2	64	128
SM	3	9(Towel), 21(Shirt), 15(Trouser)

Figure 7. Network architecture for recognition CNN.

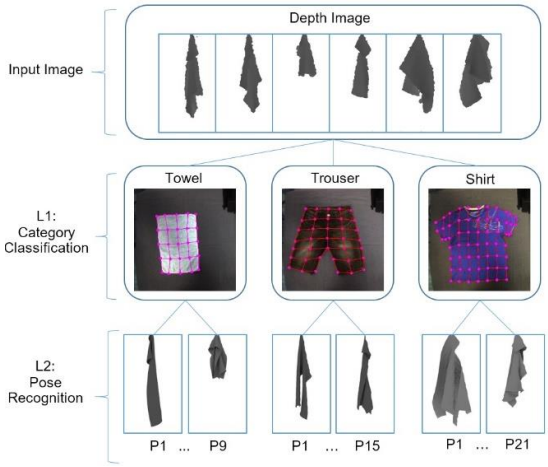
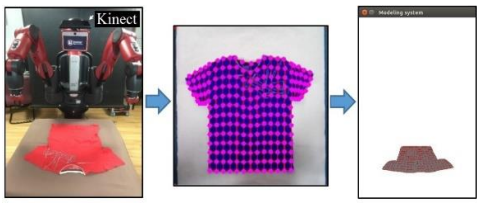


Figure 8. Overview of the proposed hierarchical approach. In the first layer the hung garment is recognized by a deep CNN. Then, a category specific CNN is performing pose estimation.

The presented CNN perform single-view classification for both category and pose recognition. However, the acquisition setup allows aggregation of the single-view results for the entire dataset of the 360 depth images that correspond to the same pose. A simple but effective method to achieve this is to perform majority voting between the different outputs of the 360 single-view classifications. This approach can be applied to both category and pose classifiers, boosting the performance and introducing more robustness to the classifiers.



(a) Place flat a garment on flat surface

(b) Generate garment mesh model

(c) Model in offline simulator

Figure 9. Experiment setup and preparation for hanging the spread-out garment mesh model before simulation.

IV. Experimental Results

A. Real and Synthetic Dataset

To collect the real garment dataset, a piece of the garment placed on the flat platform is picked up by a Baxter robot and a Kinect depth sensor located in front of the Baxter robot. The hung garment is then rotated 360 degrees vertically and the color and depth images are captured. The depth images are threshold-ed in order to obtain the garment from the background. The 200×340 pixels depth images that only consists of a piece of hanging garment are extracted from the 640×480 original images. For category classification, three categories namely towels, shirts, and trousers are selected in this work. Different views with same pose of the garment are collected for each specified grasping point. For pose recognition, 21 poses for shirts, 15 pose for trousers and 9 poses for towels are selected from the predefined points. Total 48600 depth images including towels, shirts, and trousers are

collected from Alternative 1, the procedure of image acquisition is considered slow and time-consuming.

Figure 10. Annotation points for hanging the spread-out garment mesh model before image acquisition.

Figure 11. Example images of color, Alternative 1, 2 and 3 respectively.

For the synthetic dataset, a large dataset of depth images is constructed by using Alternative 2 and Alternative 3 respectively. For Alternative 2, the total 9 garment models which 3 models of towels, 3 models of trousers and 3 models shirts are constructed with different in size and shape in Blender software. For simplify task in designing and model simulation, the garment models are constructed in 2D by assuming the front and back sides of shirt and trouser are not separated during grasped by robotic arm. In addition, the 2D garment model can also be generated by using Alternative 3 as shown in Figure 9. The particle-based polygonal model can be generated from a crudely spread-out garment placed on the flat platform. The towel models consist of 81 particles, the trouser models consist of 121 particles and the shirt model consists of 209 particles are constructed as shown in Figure 10. But only selected particles as shown in Figure 4 are selected in offline simulation. For synthetic dataset, the condition in the virtual environment is set as close as possible to the depth images in real dataset. Different views of the hung garment model can be acquired by rotating the model 360 degree in front of virtual camera that placed at fixed height as similar to Alternative 1. The total 97200 synthetic images are collected from offline simulation. As shown in Figure 11, the synthetic images acquired from the virtual simulation are comparable with real depth images which able to employ as training dataset in the CNNs later.

B. Results and Discussions

The implementation and evaluation of the convolutional neural network for category and pose recognition of garment is built with the Keras deep learning framework. Using the real and synthetic dataset for training in CNN, the accuracy and loss model for the classification of the category of garment are depicted in Figure 12. The accuracy and loss model for pose of towel are depicted in Figure 13. The accuracy and loss model for pose of shirt are depicted in Figure 14. The accuracy and loss model for pose of trouser are depicted in Figure 15. The recognition rate of 89.58% has been reported for category classification of garment. For pose recognition, the 94.31%, 88.75% and 90.83% are reported for towel, shirt and trouser respectively. From the result obtained, the recognition rate of category classification is affected by several poses of the garment. Furthermore, the dataset of garment categories is generated by collecting the hung garment model in different poses, sizes, and types. The poses of hung garment model sometimes are difficult differentiated from the captured images although they are different types. By using similar architecture of CNN, the pose recognition of towel is quite stable and accurate since the number of poses is less than shirt and trouser. From the observation, the results of the pose recognition of the garment may also affected by the inner distance between the robotic grasping points (or the selected particles in the garment model). The accuracy will be affected in learning model if the selected poses of hung garment model are close to its others. Overall, the dataset that generated from real and synthetic images by using the proposed alternatives is workable, the accuracy obtained from the training model is acceptable.

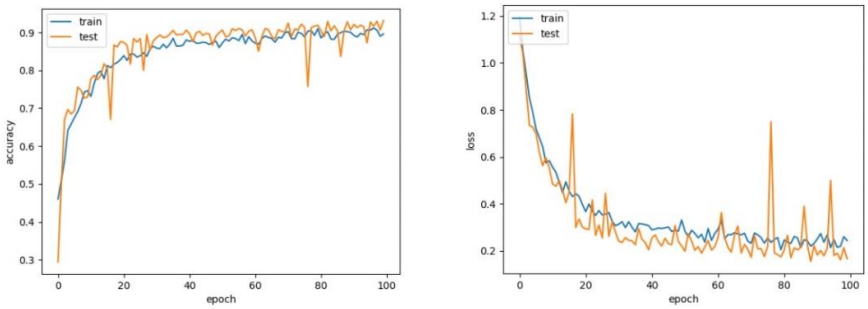


Figure 12. Training accuracy (left) and loss (right) for the category of garment.

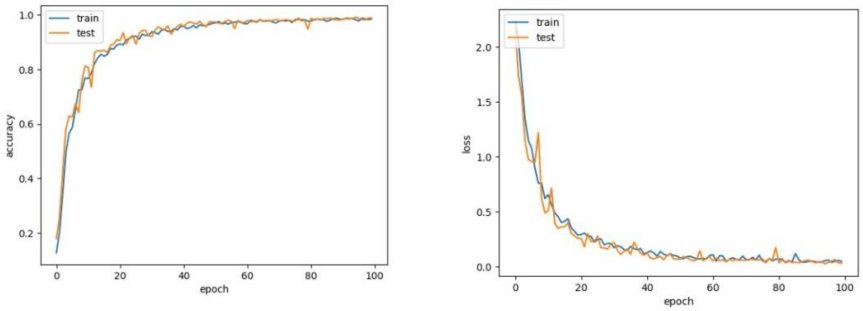


Figure 13. Training accuracy (left) and loss (right) for the towel model.

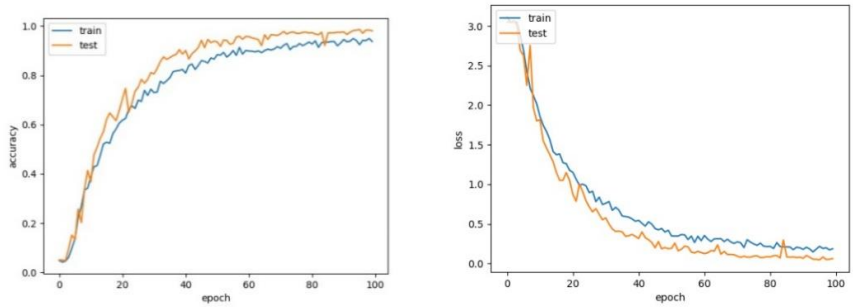


Figure 14. Training accuracy (left) and loss (right) for the shirt model.

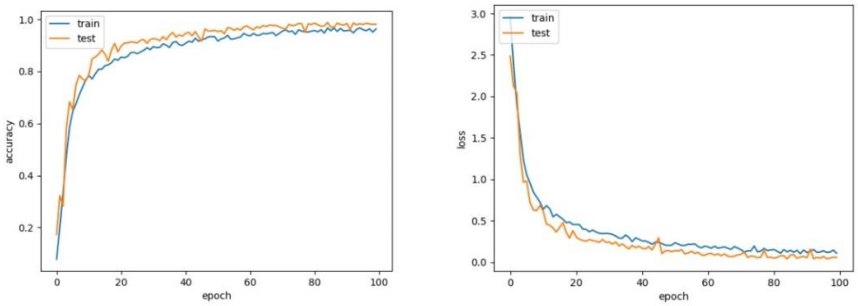


Figure 15. Training accuracy (left) and loss (right) for the trouser model.

V. CONCLUSION

The hierarchical convolutional neural networks are used to recognize the category and pose of a hanging garment. Using a low-cost depth sensor such as the Kinect, the depth images of hung garments are used as the inputs to the CNNs. Three alternatives to collect the images of hung garment are discussed. By using this proposed approach, different types and sizes of garment mesh model can be extracted and modeled in virtual simulation. The depth images from real garment and the synthetic images from garment mesh model are integrated into the dataset and then employed in the CNNs. Experiment results demonstrate the proposed approach is applicable and effective in dataset collection for CNNs that need a large volume of data to train the learning model. For future improvement, more complex and different category of garment such as long-sleeved shirts and socks will be considered and evaluate its performance once the datasets are extended. Thereafter, further works such as re-grasping and unfolding a piece of garment can be considered in the garment handling process. In addition, the proposed approach may also benefit to the pipeline of dexterous garment folding in robotic system which real garment can be represented by the garment mesh model and the robotic grasping point can be defined from the model.

ACKNOWLEDGMENT

This work was supported by the Universiti Tenaga Nasional Innovation & Research Management Centre grant number J510050872 and the Ministry of Higher Education, Malaysia through research grant: 20140127/FRGS/V3500. The authors acknowledge the CAMARO Research Group of Universiti Tenaga Nasional for the facilities and equipment.

REFERENCES

[1] Y. C. Hou, K. S. Mohamed Sahari, L. Y. Weng, D. N. T. How, and H. Seki, “Particle-based perception of garment folding for robotic

manipulation purposes," International Journal of Advanced Robotic Systems, vol. 14, no. 6, p.1729881417738727, 2017.

[2] Y. C. Hou and K. S. M. Sahari, "Self-generated dataset for category and pose estimation of deformable object," Journal of Robotics, Networking and Artificial Life, vol. 5, no. 4, pp. 217–222, 2019.

[3] F. Osawa, H. Seki, and Y. Kamiya, "Unfolding of massive laundry and classification types by dual manipulator," JACIII, vol. 11, no. 5, pp. 457–463, 2007.

[4] K. S. M. Sahari, H. Seki, Y. Kamiya, and M. Hikizu, "Real-time path planning tracing of deformable object by robot," International Journal on Smart Sensing & Intelligent Systems, vol. 3, no. 3, 2010.

[5] 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.

[6] S. Miller, J. Van Den Berg, M. Fritz, T. Darrell, K. Goldberg, and P. Abbeel, "A geometric approach to robotic laundry folding," The International Journal of Robotics Research, vol. 31, no. 2, pp. 249–267, 2012.

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

[8] D. Triantafyllou, I. Mariolis, A. Kargakos, S. Malassiotis, and N. Aspragathos, "A geometric approach to robotic unfolding of garments," Robotics and Autonomous Systems, vol. 75, pp. 233–243, 2016.

[9] Y. Kita, T. Ueshiba, E. S. Neo, and N. Kita, "Clothes state recognition using 3d observed data," in IEEE International Conference on Robotics and Automation (ICRA), pp. 1220–1225, IEEE, 2009.

[10] A. Doumanoglou, A. Kargakos, T.-K. Kim, and S. Malassiotis, "Autonomous active recognition and unfolding of clothes using random decision forests and probabilistic planning," in Robotics and Automation (ICRA), 2014 IEEE International Conference on, pp. 987–993, IEEE, 2014.

[11] A. Doumanoglou, J. Stria, G. Peleka, I. Mariolis, V. Petrik, A. Kargakos, L. Wagner, V. Hlav´a˘c, T.-K. Kim, and S. Malassiotis, "Folding clothes autonomously: A complete pipeline," IEEE Transactions on Robotics, vol. 32, no. 6, pp. 1461–1478, 2016.

[12] I. Mariolis, G. Peleka, A. Kargakos, and S. Malassiotis, "Pose and category recognition of highly deformable objects using deep learning," in Advanced Robotics (ICAR), 2015 International Conference on, pp. 655–662, IEEE, 2015.

[13] Y. Li, Y. Wang, Y. Yue, D. Xu, M. Case, S.-F. Chang, E. Grinspun, and P. K. Allen, "Model-driven feedforward prediction for manipulation of deformable objects," IEEE Transactions on Automation Science and Engineering, no. 99, pp. 1–18, 2018.

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

[15] Y. Li, Y. Wang, M. Case, S.-F. Chang, and P. K. Allen, "Real-time pose estimation of deformable objects using a volumetric approach," in Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on, pp. 1046–1052, IEEE, 2014.

[16] E. Corona, G. Alenya, A. Gabas, and C. Torras, "Active garment recognition and target grasping point detection using deep learning," Pattern Recognition, vol. 74, pp. 629–641, 2018.