In-Hand Manipulation of Multi-Fingered Hand with Daily Objects Based on Graph Convolutional Network and 3-Axis Tactile Sensing

Satoshi Funabashi [1], Tomoki Isobe [1], Fei Hongyi [1], Tetsuya Ogata [2],
Alexander Schmitz [1] and Shigeki Sugano [1]

***Abstract*— Multi-fingered hands have potential for many applications where dexterous manipulation as humans do is required. To enhance its manipulation stability with a variety of objects, tactile sensing is important. However, tactile sensors on the multi-fingered hands are mounted in a variety of sizes and shapes, thus how to process such abundant tactile information and utilize it for controlling the hands is still an open issue. This paper presents a control method based on a graph convolutional network (GCN) which extracts geodesical features of the tactile information from complicated sensor alignments. Moreover, object property labels are embedded to the GCN to adjust in-hand manipulation motions. Distributed tri-axial tactile sensors are mounted on fingertips, finger phalanges and a palm resulting in 1152 tactile measurements. Training data is collected by a data-glove to transfer humans dexterous manipulation directly. As a result, the GCN extract tactile features and achieved the highest success rate of in-hand manipulation. Also, object labels enabled the GCN to adjust motions resulting in achieving manipulation with a deformable object without squeezing it.**

I. INTRODUCTION

Humans use their multi-fingered hands for dexterous manipulation. Fingers moving in synchrony realize various motions such as rolling and finger gating. Furthermore, human skin supports those manipulation stability. To realize the manipulation by robotic hands, those finger motions and tactile sensing skills are essential, otherwise this is not possible because occlusion by the hands easily happens. Specifically, 3-axis tactile sensing is useful as multiple fingers touch an object at the same time from different orientation during manipulation. Also, when the hands grasp the objects, grasping postures are changed by the size and shape of the objects and thus contact positions on the hands are diverse such as fingertips, finger phalanges and a palm. For this reason, multi-fingered hands should have distributed 3-axis tactile sensors on the surface of the hands as much as possible. For manipulating a variety of objects, object properties such as slipperiness and softness also need to be considered otherwise the hands can drop or squeeze the objects. Therefore, tactile sensors play another important role for recognizing such properties. Overall, if a controller of the hands recognizes the properties and change a manipulation motion, the hands can achieve in-hand manipulation with various objects (e.g. the controller recognizes that an object is soft and generates a gentle motion resulting in successful manipulation without breaking the object).

[1]Satoshi Funabashi, Tomoki Isobe, Fei Hongyi, Alexander Schmitz and Shigeki Sugano are with the Faculty of Science and Engineering, Dept. of Modern Mechanical Engineering, Waseda University, Tokyo 169-8555, Japan. (e-mail: s_funabashi@sugano.mech.waseda.ac.jp, t_isobe@sugano.mech.waseda.ac.jp, feihongyi@asagi.waseda.jp, schmitz@aoni.waseda.jp, tito@toki.waseda.jp, sugano@waseda.jp)

[2]Tetsuya Ogata is with the Faculty of Science and Engineering, Dept. of Intermedia Art and Science, Waseda University, Tokyo 169-8555, Japan. (e-mail: ogata@waseda.jp)

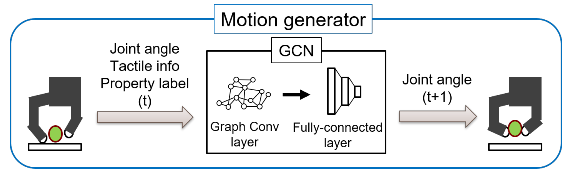

Fig. 1: Schematic of the proposed motion-generating method.

Graph convolutional network (GCN) is focused in this study. The GCN is used for graph-structured applications such as molecules and traffic networks. This study investigates the GCN applied to tactile sensor alignments as following robotic configuration geodesically without any experimenter's engineering of tactile information. When increasing the number of convolution layers, each node is convoluted with surrounding nodes and that makes the network acquire tactile features geodesically. This can be utilized for recognizing whole contact states on each part of a robotic hand resulting in successful fingers motions in synchrony.

II. RELATED WORK

Tactile sensors enable robotic hands to do dexterous manipulation. Specifically, tactile sensors for fingertips such as Biotac [1] and [2] GelSight are widely used. Using these tactile sensors is beneficial for achieving not only manipulation but also recognition of object or object property.

By adding tactile information to control methods, the hand can be more dexterous and handle a variety of objects. However, there were not suitable tactile sensors which fit to multi-fingered hands. Specifically, tactile sensors should have tri-axial and distributed tactile information to detect events on the hands. Also, the sensors should be able to be bent for fingertips. Finally, uSkin, distributed 3-axis tactile sensors were mounted on a multi-fingered hand namely Allegro Hand in our previous research [3]. CNNs are widely used for tactile based robotic tasks these days. However, the network requires to have input in a rectangular shape, and thus tactile information needs to be reshaped by an experimenter resulting in unstable results.

On the other hand, a study used GCN [4] for Biotac sensor which has unstructured tactile sensor alignments [5].

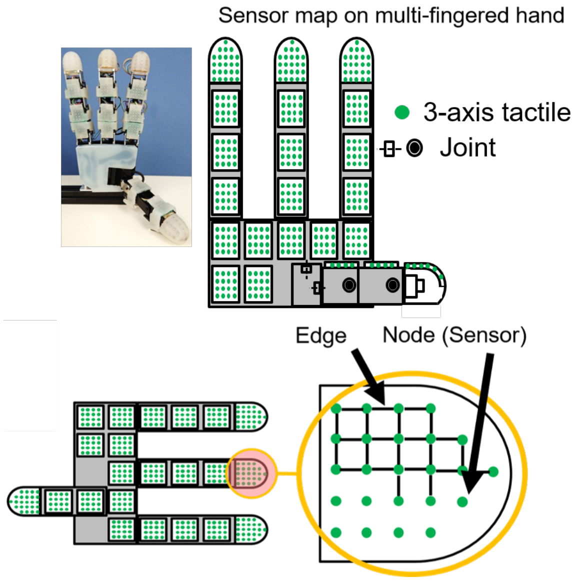

Fig. 2: Tactile sensors and graph structures

However, the GCN was only applied to an area on fingertips.The GCN can also be applied to tactile sensors by a multi-fingered hand.

III. PROPOSED METHOD

An Allegro Hand with uSkin sensors is used. Each sensor point of the uSkin as a node, are connected by an edge as a graph structure (Fig. 2). By constructing the information of each node and edge together with GCN, the grasping state of an object can be recognized with high accuracy. In addition, by introducing GCN into the control method of the hand, we can stabilize the results because there is no need to reform the tactile sensor map as for CNNs realizing highly accurate object recognition. In addition, it is possible to input all sensors together without losing sensor positional information even in complex sensor arrangements.

A model schematic of the motion generator used in this study is shown in Fig. 1. When the robot hand starts manipulation from the initial grasping posture, the GCN receives sensor information. Specifically, the tactile sensor information is first input to the graph convolution layer. Then, the features obtained by the output from the graph convolution layer, the joint angle and property labels are input to all the fully connected layers. Next timestep of joint angles are output to adjust the posture of the fingers. By repeating these series of generation, the final grasping posture is reached.

IV. EXPERIMENT DESIGN

A. Training Data

We selected the movement of grabbing an object from the floor as the target movement. The reasons for selecting this motion was that it makes contacts with the entire hand and hence tactile and geodesical information are more important. Each recorded training data is pre-processed before being input to the GCN. First, in each recorded training data, the part where the finger is not moving immediately after the start of recording and the part where the finger is not moving after the end of manipulation are cut out. Then, label information corresponding to the object used to acquire each training data is appended to the training data. Specifically, we prepare six labels in the following order: light, heavy, hard, soft, non-slippery, and slippery, and fill in 1 for each label if it applies, and 0 if it does not. For example, for a heavy, soft, and non-slippery object, the labels would be [0,1,0,1,1,0].

B. Neural Network Settings

First of all, in order to verify the effectiveness of GCN in this research, we conducted comparative experiments when the neural network in the motion generator was GCN and when it was multi-layer perceptron (MLP). Manipulation with each model was conducted five times. When the network was GCN, the experimental setup was six graph convolution layers with sizes of [14, 28, 56, 112, 112, 112], no pooling layer, and four total fully-connected layers with sizes of [8000, 1000, 120, 50]. The input for the GCN was tactile, joint measurements and object property labels. The output was a next timestep of the joint measurements. The number of dimensions of the input is 4 (finger) x 4 (joint) = 16 dimensions for joint angle, 384 (sensor) x 3 (axis) = 1152 dimensions for tactile information, and 6 dimensions for object property, so the total number of dimensions of the input is 1174. The number of dimensions of the output is 4 (fingers) $\times$ 4 (joints) = 16 dimensions. The number of timesteps used in this experiment was 340 and the total timestep data used in this experiment was 26300, of which the timestep data used for training is 18410, and for testing is 7890. The learning rate of the Adam optimizer is 0.00001. The batch size is 100 for both training and testing. On the other hand, when the network is MLP, the experimental setup is as follows: the total number of fully-connected layers is 7, and the size of each layer is [1500, 3000, 1500, 700, 350, 100, 50], there is no pooling layer. The input is tactile, joint angle and object property information. The output is joint angle. The number of dimensions of the input and output are the same as those of GCN.

V. EVALUATION

A. Comparison of Neural Networks

We used a hard plastic cylinder and a soft plastic cylinder in this comparison experiment. The Allegro Hand using the model with GCN (model I) and MLP (model IV) were controlled. For the GCN, the joint angles of the fingers were finely adjusted to manipulate the object (5 times out of 5 trials), and the movements of the index, middle, and ring fingers were different. On the other hand, the final grasping posture of the MLP was different from that of the training data, and the object could not be lifted to a sufficient height. Manipulation was successful in one out of five trials. When the object was a soft plastic cylinder, the GCN succeeded in manipulating it 3 times out of 5 trials, and the MLP succeeded in manipulating it 0 times out of 5 trials. When the

TABLE I: Achievement of the Final Posture with hard plastic cylinder

Model	Success Rate
I	5/5
II	1/5
III	1/5
IV	0/5

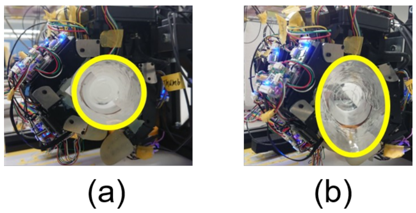

Fig. 3: Grasping states with a soft plastic tube. (a) shows the final grasping posture with a model given the correct labels. The grasped object is not deformed from its side. (b) shows the final grasping posture with a model given the wrong labels. The grasped object is deformed in oval shape from its side.

object was a sponge, the GCN succeeded 3 out of 5 times, and the MLP succeeded 0 out of 5 times.

In order to investigate the changes in the manipulation motion due to differences in the structure of the GCN, networks with three different graph convolutional layer structures were built. They performed five manipulation trials. The three models are referred to as Model I, Model II, and Model III, respectively. Model I has six graph convolution layers with sizes of [14,28,56,112,112,112]. In Model II, the number of layers of the graph convolution layer is 4 and the size is [14,28,56,112], and in Model III, the number of layers of the graph convolution layer is 3 and the size is [14,28,56]. Also, the common parameter settings for the three models are no pooling layer, four fully-connected layers, and size of [8000,1000,120,50]. The manipulation using Model II often failed because the distance between the palm and the object was more than 2cm in the final grasping posture, and the manipulation succeeded only once out of five trials. The same result happened with Model III.

B. Analysis on Touch States with Labels

Next, in order to confirm the effectiveness of using property labels, we performed manipulation by changing the property labels during in-hand manipulation. In this comparison experiment, the property labels about the grasping object was specified and input to the GCN (model I) during the manipulation. A soft plastic tube was selected as the object in this comparison experiment. Two types of property labels were used as input: correct labels and incorrect labels. The correct labels consist of light, soft, and slippery, which are the properties of the object in this experiment, while the wrong labels consist of heavy, hard, and slippery. Fig. 3 shows the manipulation and the cross-section of the object in the final grasping posture when the correct property label (Fig. 3-(a)) and the incorrect property label (Fig. 3-(b)) are

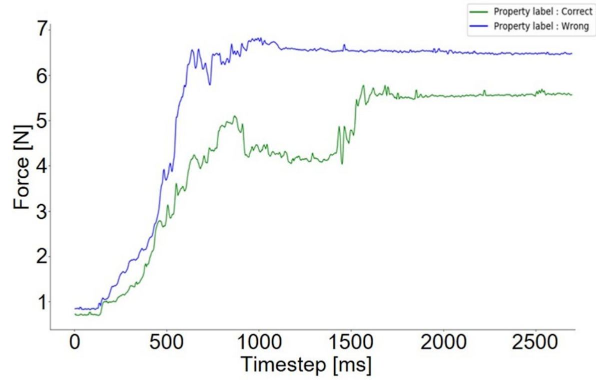

Fig. 4: Trajectories of grasping forces on a plastic tube are shown. Blue line shows the forces generated by a model given wrong labels. Green line shows the forces generated by a model given correct labels.

used as input. When the correct property label was used, the soft plastic cylinder did not collapse, and the cross section was circular as shown by the yellow line. On the other hand, when the wrong property label was used, the soft plastic cylinder was crushed, and the cross section was deformed into an oval as shown by the yellow line.

As shown in Fig. 4, the total force at each tactile sensor was always higher when the wrong property label was used than when the correct property label was used. From this result, it can be said that when the correct label was used, the robot is able to perform the manipulation with an appropriate grasping force, and therefore, does not crush the object.

VI. CONCLUSIONS

This study showed a control method that a multi-fingered hand manipulates a daily object. A GCN acquiring tactile and geodesical features of a robot hand achieved dexterous in-hand manipulation of synchronized fingers. Furthermore, labels for each object property enabled the GCN to change manipulating motions depending on the target object and grasping forces were reduced.

As future works, the GCN would be improved to achieve more stable in-hand manipulation with a variety of objects.

REFERENCES

[1] F. Veiga, H. van Hoof, J. Peters, and T. Hermans, "Stabilizing novel objects by learning to predict tactile slip," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Sep. 2015, pp. 5065–5072.

[2] W. Yuan, C. Zhu, A. Owens, M. A. Srinivasan, and E. H. Adelson, "Shape-independent hardness estimation using deep learning and a gelsight tactile sensor," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, May 2017, pp. 951–958.

[3] S. Funabashi, S. Morikuni, A. Geier, A. Schmitz, S. Ogasa, T. P. Tomo, S. Somlor, and S. Sugano, "Object recognition through active sensing using a multi-fingered robot hand with 3d tactile sensors," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Sep 2018 (Accepted).

[4] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2017.

[5] A. Garcia-Garcia, B. S. Zapata-Impata, S. Orts-Escolano, P. Gil, and J. Garcia-Rodriguez, "Tactilegcn: A graph convolutional network for predicting grasp stability with tactile sensors," in *2019 International Joint Conference on Neural Networks (IJCNN)*, 2019, pp. 1–8.