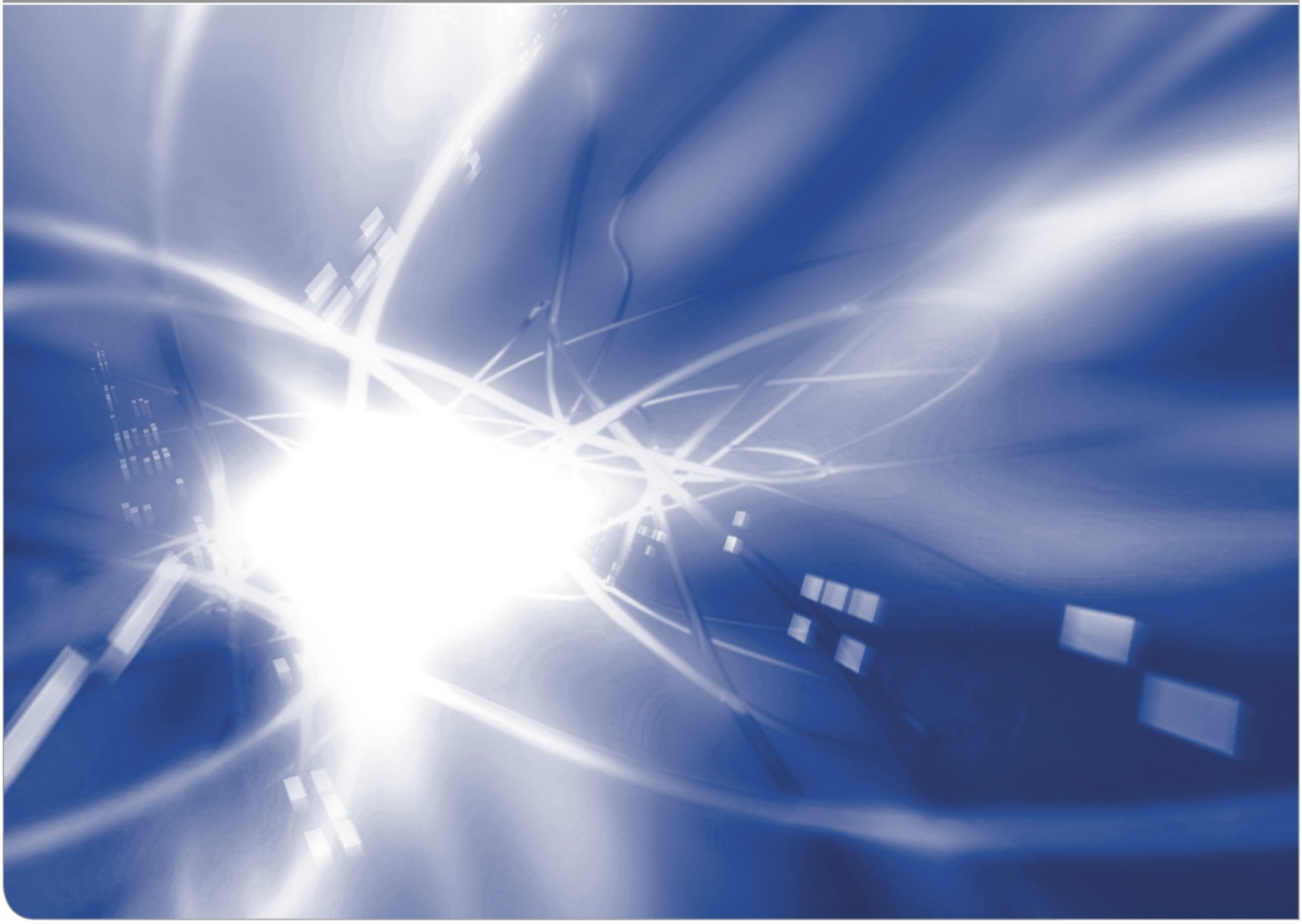
Karlsruhe Institute of Technology

Proceedings of the 5th Workshop on Proximity Perception in Robotics at IROS 2022: Towards Next-Generation Multi-Modal Sensing in Soft Structures

Edited by

Hosam Alagi[1], Natalia Sanchez-Tamayo[2], Yitao Ding[3], Christian Schöffmann[4], Zackory Erickson[5], Kaspar Althoefer[6], Hubert Zangl[4], Harald Gietler[4], Björn Hein[1,7]

October 23th, 2022, Kyoto, Japan.
IEEE/RSJ International Conference on Intelligent Robots and Systems



Introduction

The 5th "Workshop on Proximity Perception in Robotics" was held in Kyoto, Japan on October 23rd, 2022 as part of the 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). This workshop version continued the topic of multi-modal perception and expanded the scope to open research questions in sensor design and modeling of soft structures, as well as the development of realistic simulation models for applications in robot learning and AI.

The workshop focused on (i) new technological developments of proximity and tactile sensing systems embedded in soft structures, (ii) realistic modeling and simulation of such sensing systems, and (iii) state-of-the-art applications in robot learning and manipulation. Industry speakers provided first-hand accounts of the procedures for certifying and deploying proximity and tactile perception technologies on real robotic systems. Speakers from academia provided expertise in cutting-edge proximity and tactile perception systems, along with innovative solutions for their applications. In addition, the workshop featured a PhD forum session and paper presentations to disseminate early research to a broad interdisciplinary audience. As a result, the workshop continued to build connections between academic research and industry as well as between the communities of proximity perception, tactile perception, and emerging robotic materials.

These proceedings contain the submissions accepted to the workshop and presented during the PhD-Forum.

Talks held at the workshop:

Human-Robot Interaction
Oliver Brock
Technical University of Berlin, Germany

Model-Based Sensing for Soft Robots
Stefan Escaida Navarro
Universidad de O'Higgins, Chile / Inria Lille

Robots working with and around people
Alessandro Roncone
University of Colorado Boulder, USA

Enriching humanoid robot interactions with large-area e-skin
Gordon Cheng (represented by Rogelio Guadarrama)
Technical University of Munich, Germany

Energy Generating Large Area Electronic Skin
Ravinder Dahiya
University of Glasgow, UK

Human-Robot Interaction in Practice
David Reger
Neura Robotics GmbH, Germany

For more information, please visit the workshop website.
proxelsandtaxels.org

Vision-Guided Pre-Shaping from Tactile Memory for In-Hand Manipulation

Satoshi Funabashi Fei Hongyi Alexander Schmitz Shigeki Sugano and Tetsuya Ogata

Abstract— **Multi-fingered robotic hands can achieve various tasks like a human. So far, dexterous manipulation has been achieved using Deep Learning. However, preshaping is important during reaching an object from a distance to the object to start optimal in-hand manipulation. In this case, it is necessary to recognize the external and internal properties of the object. Therefore, this study aims to achieve stable success in generating pre-shaping motions that takes into account the grasping position and object properties. Specifically, training was performed via a motion generator, and a combination of Mask R-CNN for processing image information and convolutional autoenconder (CAE) and long-term short-term memory (LSTM) for generating the motions amd tactile memory from image information. In an ablation study, the model with contoured information and tactile memory achieved a high success rate of generating pre-shaping motions.**

I. INTRODUCTION

In recent years, the use of neural networks to achieve robotic manipulation has been a very popular research direction in recent years. Many of them use vision input information. However, in the case of multi-fingered robot hands, the visual information is subject to occlusion, which makes achieving robotic manipulation by neural networks significantly more difficult. For this reason, tactile information is expected to be useful for achieving such manipulation, We also proposed a graph convolutional network (GCN) as a motion generator and the network learnt the positional relationships between the individual sensors [1]. The GCN successfully manipulated objects with different properties. However, since the initial grasping posture was the same for all objects, it was not possible to change the initial grasping posture (i,e, pre-shaping) depending on the objects' properties. As a result, the manipulation motion often became unstable when the object was changed. Due to the lack of a combination of visual and tactile information, it is also difficult for reaching the target objects.

Preshaping was introduced by Stansfield et al.[2] in 1988 as one stage of object manipulation behavior with a robot hand. Recent work on preshaping was done by Liu et al. in 2019 [3], who were able to generate reasonable preshaping postures for objects with various external shapes via CNN. However, this study generated many objects in simulation, and for some objects it is difficult to generate the motion as it would be in reality if the object were to contact an environment (e.g., a desk). In [4], objects with the same shape change humans' preshaping postures depending on their material, or internal properties. Therefore, we considered that not only the external properties of the target objects but also its internal properties were important when generating a preshaping motion. For how to combine visual tactile information, [5] successfully classified the internal properties of an object by the latent variables of a neural network by relating the tactile information and the visual information using the distributed 3-axis tactile sensors. Specifically, by using the visual information an image of an object as the input information for the encoder and the tactile information when touching an object as the output information for the decoder, the neural network was trained to classify the internal properties of an object in the latent variables between the encoder and decoder. From those studies, the objective of this study was to use visual information and tactile memory to generate pre-shaping motions based on the external and internal properties of an object to stabilize the manipulation behavior.

The authors are with Waseda University, Okubo 3-4-1, Shinjuku, Tokyo 169-8555, Japan. (e-mail: s-funabashi@aoni.waseda.jp).

II. PROPOSED METHOD

In this study, visual information was processed by Mask R-CNN [6] to generate the outer contour of the corresponding object. The visual data was then passed to CAE-LSTM by compressing the data with the CAE part. The LSTM predicts the tactile information that matches the properties of the object as tactile memory. The LSTM also predicts the joint angles of the robot hand and arm and image features for the next timestep, and the joint angles were used to control the robot hand and arm for generating pre-shaping motions.

III. EXPERIMENT DESIGN

The overall system of this experiment is shown in Fig. 1, with the D435 camera fixed directly above the experimental table, and the robot system is consisted of the Torobo Arm and Allegro Hand. In addition, to facilitate the processing of visual information, the experimental platform is covered with green artificial grass.

In this study, objects with various properties were selected as target objects. This study focused on two properties, size and hardness, and prepared a total of eight types of objects combining each characteristic of large/small and hard/soft. Specifically, Potato chip tube (large and hard), Saran wrap (large and hard), Large sponge (large and soft), Kitchen paper (large and soft), Dice (small and hard), Stuffed melon bread

5th Workshop on Proximity Perception
2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2022), KYOTO, JAPAN

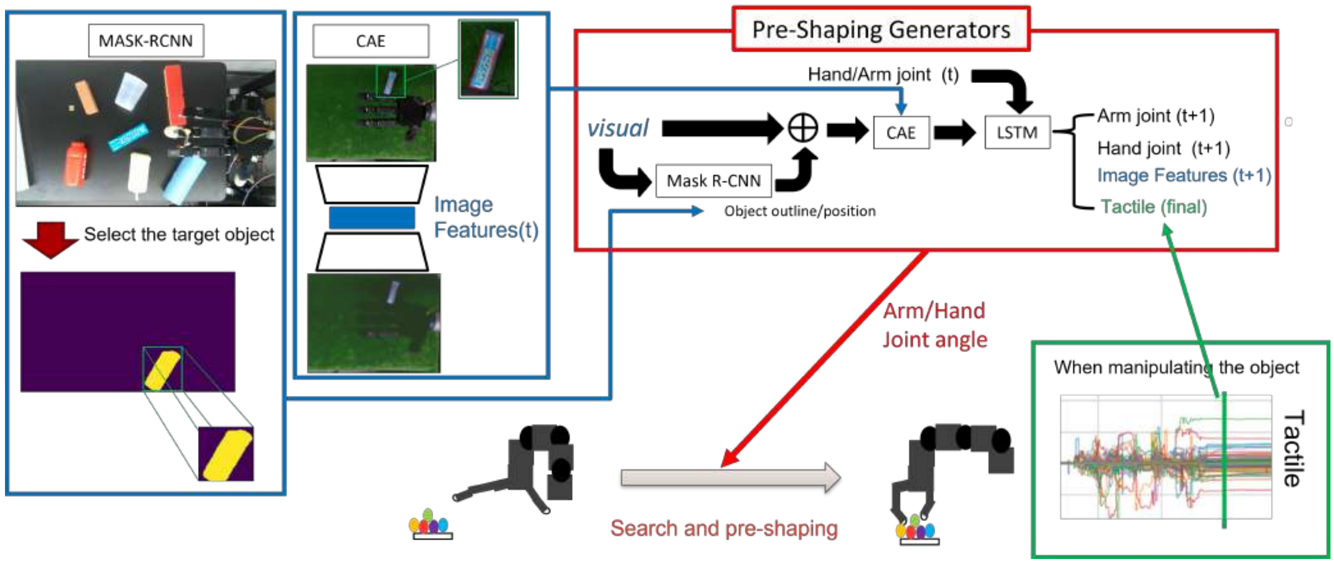


Fig. 1. Proposed network architecture using Mask-RCNN (vision) and CAE-LSTM (motion generator with tactile memory).

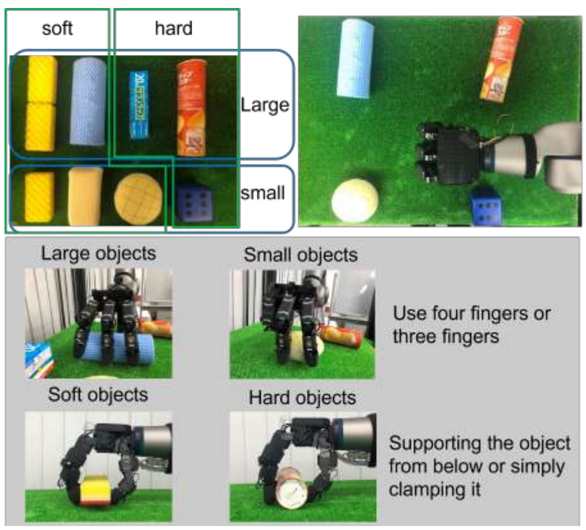


Fig. 2. 8 different objects based on softness and size are chosen. Four object positions are chosen as reaching positions. Preshaping posture is prepared for each object property.

(small and soft), Sponge (small and soft) and Small sponge (small and soft) were selected.

In the evaluation experiment of this study, we selected a preshaping motion toward a target object placed at a random position on a desk among several objects. The reasons for selecting this motion were (i) the object needs to be pinched with a precision grasp before performing in-hand manipulation, (ii) object position information was necessary because there are several objects and they are randomly located and (iii) the objects had various internal properties that needed to be predicted when an appropriate preshaping posture was needed. In order to confirm that the proposed system changes the grasping posture depending on the internal and external properties of the object, the system uses all four fingers for large objects and only three fingers for small objects, as shown in Fig. 2. For a soft object, the posture is as if supporting the object from below, and for a hard object, the posture is as if simply pinching the object. For collecting training data, the Torobo arm was set to move to the grasping points while the Allegro Hand was remotely controlled using Cyberglove. A total of four data collection points were set as shown in Fig. 2. Each trial was collected for 15s at 100Hz. Two trials for each object was collected. A total of 64 trials was collected with each of the four data collection points, Each training data was preprocessed before being input to the proposed networks. Firstly, only tactile information that was in contact with the target object was used as tactile memories. Each motion data was divided into 10 dataset divided by 10 timesteps in each trial, resulting in a total of 576 (8 objects x 8 trials x 9) sets of inputs. This study aims to generate tactile memories by correlating visual and tactile information. To achieve this, preprocessing of the training data is necessary. Since the tactile data before the contact between the robot hand and the object is meaningless, we used only the tactile data from the stable situation after the contact between the robot hand and the object, i.e., the last time step of the entire movement, as shown in Fig. 1.

IV. EVALUATION

First, to verify the effectiveness of Mask R-CNN in this study, a comparison experiment was conducted with and without Mask R-CNN in the proposed model for object segmentation, and five trials were attempted with each model. In the experiment, two types of objects (sponge and kitchen paper) and two types of untrained objects (plastic bottle case and candy) were placed side by side. Fig. 3 and Fig. 4 shows when the pre-shaping motion was for a kitchen paper.

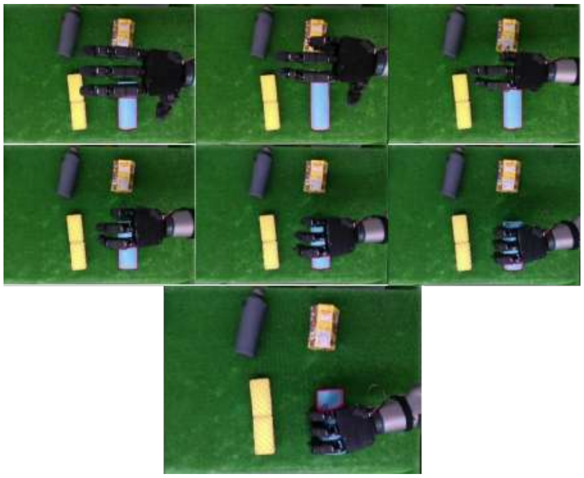


Fig. 3. Tactile sensor alignments and its graph structure. (a) shows that uSkin sensors are mounted on the fingertips, phalanges and palm. (b) shows how we built the graph structure of tactile sensors. Each sensor chip is regarded as a node and they are connected by edges.

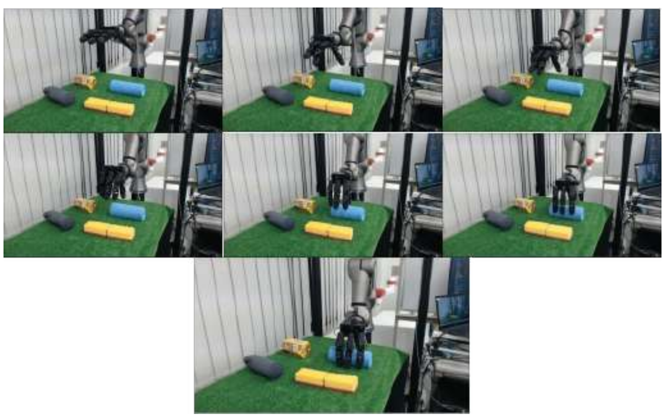


Fig. 4. Tactile sensor alignments and its graph structure. (a) shows that uSkin sensors are mounted on the fingertips, phalanges and palm. (b) shows how we built the graph structure of tactile sensors. Each sensor chip is regarded as a node and they are connected by edges.

It was confirmed that the hand was positioned correctly. The experiment with this object was successful 5 out of 5 times. Without segmentation with Mask R-CNN, the model firstly approached the kitchen paper and switched to the sponge after some timesteps. Therefore, the model was never successful in all five experiments.

Second, whether using tactile memory was useful or not was confirmed. The model using tactile memory achieved a high success rate (more than three successful experiments for all target objects). On the other hand, the model without tactile memory performed 5 experiments on each of the objects tested, with 2 objects succeeding 2 out of 5 times and all other objects failed. From this result, the use of tactile memory was confirmed to be useful for preshaping.

V. CONCLUSIONS

This study proposed a neural network-based pre-shaping motion generator that processes visual information with Mask R-CNN and CAE, and generates motion and tactile memory by linking visual and tactile information with LSTM. In the evaluation experiment of this study, we selected a pre-shaping motion of a target object from a random position on the floor among several objects. Comparison experiments were conducted between models using Mask R-CNN and models without Mask R-CNN. In this experiment, the model with Mask R-CNN succeeded in automatically preshaping the target object. The model with tactile memory achieved a high success rate of preshaping motions. From these results, using Mask R-CNN and tactile memory was confirmed to be useful for achieving preshaping motions depending on object properties.

REFERENCES

[1] S. Funabashi, T. Isobe, F. Hongyi, A. Hiramoto, A. Schmitz, S. Sugano, and T. Ogata, “Multi-fingered in-hand manipulation with various object properties using graph convolutional networks and distributed tactile sensors,” *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 2102–2109, 2022.

[2] S. S., A., “Reasoning about grasping,” in *1988 National Conference on Artificial Intelligence*, 1988.

[3] M. Liu, Z. Pan, K. Xu, K. Ganguly, and D. Manocha, “Generating grasp poses for a high-dof gripper using neural networks,” 2019. [Online]. Available: https://arxiv.org/abs/1903.00425

[4] V. C. Paulun, K. R. Gegenfurtner, M. A. Goodale, and R. W. Fleming, “Effects of material properties and object orientation on precision grip kinematics,” 2016. [Online]. Available: https://arxiv.org/abs/1903.00425

[5] K. Takahashi and J. Tan, “Deep visuo-tactile learning: Estimation of tactile properties from images,” in *2019 International Conference on Robotics and Automation (ICRA)*, 2019, pp. 8951–8957.

[6] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask r-cnn,” in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2980–2988.

A Practical Guide to Embedded 3D Printing for Soft Skin-like Sensors

Sonja Groß*[,1,2] Silija Breimann*[,1], Amartya Ganguly[1] and Sami Haddadin[1,2]

Abstract— **In recent years, 3D printing technology has introduced new opportunities for efficient, scalable and versatile manufacturing of soft tactile sensors. In particular, embedded 3D (e-3D) printing, which deposits functional ink directly into a polymer layer, reduces the necessity for multi-step casting and creates opportunities for improved shape fidelity and highly customizable designs. To further increase accessibility for this method, that can serve as an optimal platform for soft tactile skin developments in various applications, this work will focus on providing practical guidance about materials, pattern design, parameterization and the soft/hard interface to electronics.**

We used a commercial 3D printer combined with a high precision soft materials print head from which carbon grease was deposited into a double-layered silicone matrix to create skin-like soft contact sensors (see fig.1-a, b, c) [1]. To achieve optimal print results (see Fig. 1-d), the mechanical and chemical behaviour of the printed material needs to match the silicone matrix material to preserve the desired shape. Materials with shear thinning behaviour are preferred and the diameter of the nozzle increases with high viscosities. Successful e-3D printing was conducted with carbon conductive grease [2], silicone mixed with carbon nano particles [3] and ionogel [4]. Furthermore, the conductive ink must be deposited into the bottom silicone layer with higher viscosity to maintain its shape. Optimal results were achieved with a bottom reservoir layer of 2mm thickness, a top filler layer of 0.5 mm and a print height of 1.5 - 1.7 mm, whereby, the 0.2 mm interval is caused by manufacturing tolerances of the casting molds. Printing higher or lower, resulted in exposed or deformed sensor patterns, as shown in Fig. 1-e. Disruptions in the print pattern as shown in Fig. 1-f can be caused by a mismatched extrusion factor (4 recommended) and respective printing speed (between 100 mm/s and 200 mm/s recommended) or insufficient air pressure at the print head (between 1-2bar recommended). Another issue is material dragging as shown in Fig. 1-g. To prevent this, we reduced the printing speed at curved lines or edges (200 mm/s) and recommend a minimal distance between lines of 1 mm. As depicted in figure 1-h, we first used flexible cables of 2.5 mm diameter that were stripped, pierced through the silicone to connect to the conductive material and sealed with a drop of fast curing silicone (FAST Ecoflex 00-35, Smooth-On Inc.), making the resulting hard-soft interface poor in robustness. Therefore, we developed an interface, mechanically decoupled from cable movements; we sewed conductive yarn (Adafruit, USA, 2ply, resistivity 0.52 Ω/cm) through the sensors and stitched it to the corners of the silicone layer Fig. (1-i).

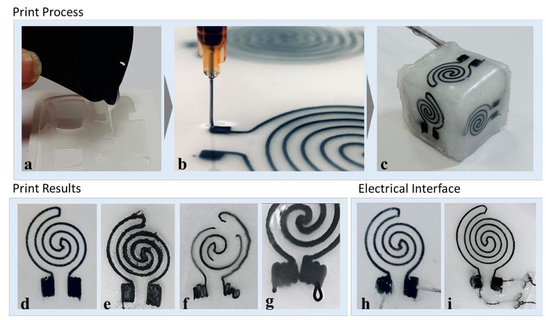


Fig. 1. Overview of the printing process and its optimization. a: preparation of the double-layered silicon matrix, b: e-3D printing. The nozzle deposits conductive ink into the matrix. c: skin-like contact sensor on a cube, d: optimal print result, e: low print height, f: disruptions, g: material dragging, h: cable interface, i: yarn interface.

This guide is intended to lower the barrier for researchers to utilize our proposed e-3D printing platform.

ACKNOWLEDGMENT

This work was funded by the German Research Foundation (DFG, Deutsche Forschungsgemeinschaft) as part of Germany's Excellence Strategy – EXC 2050/1 – Project ID 390696704 – Cluster of Excellence "Centre for Tactile Internet with Human-in-the-Loop" (CeTI) of Technische Universität Dresden. This work was also supported by the Federal Ministry of Education and Research of the Federal Republic of Germany (BMBF) by funding the project AI.D under the Project Number 16ME0539K.

REFERENCES

[1] Groß and Breimann et al., "Embedded 3d printing: A cost-effective development platform for tactile sensors," in *Haptics: Science, Technology, Applications, 13th International Conference on Human Haptic Sensing and Touch Enabled Computer Applications, EuroHaptics 2022 Hamburg, Germany, May 22–25, 2022, Proceedings*, S. et al., Ed. Springer, 2022, pp. 401–403.

[2] Muth et al., "Embedded 3d printing of strain sensors within highly stretchable elastomers," *Advanced Materials*, vol. 26, pp. 6307–6312, 2014.

[3] Wei et al., "Flexible and stretchable electronic skin with high durability and shock resistance via embedded 3d printing technology for human activity monitoring and personal healthcare," *Advanced Materials Technologies*, vol. 4, pp. 1–9, 2019.

[4] R. Truby, "Embedded three-dimensional printing of autonomous and somatosensitive soft robots," Doctoral dissertation, Graduate School of Arts & Sciences, 2018.

* Shared first authorship on this paper due to equal contribution.

[1] {sonja.gross, silija.breimann, amartya.ganguly, haddadin}@tum.de, Chair of Robotics and Systems Intelligence, MIRMI - Munich Institute of Robotics and Machine Intelligence, Technical University of Munich (TUM), Munich, Germany.

[2] and also with the Centre for Tactile Internet with Human-in-the-Loop (CeTI).

5$^{\text{th}}$ Workshop on Proximity Perception
2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2022), KYOTO, JAPAN

Learning to adapt behaviours in force-based manipulation of soft object

Marina Y. Aoyama, João Moura, Namiko Saito, and Sethu Vijayakumar

5[th] Workshop on Proximity Perception PhD Forum

I. Introduction

One of the important capabilities of fully automated robots is adaptability. For example, when grasping an object, the robot needs to adapt the strength of the grasp depending on the object's characteristics, preventing it from either breaking or slipping away.

While there is some evidence showing that robots can utilise force feedback to adapt behaviour in pouring [1] or assistive rehabilitation [2], utilising characteristics of the object or the human partner to adapt behaviour, learnt from demonstration, remains a challenge.

In this work, we investigate how to modify a Multiple Timescale Recurrent Neural Network (MTRNN) for adapting the learnt behaviour according to the estimated characteristics of the manipulated object. We evaluate our proposed modifications in the task of cutting soft objects.

II. Method

The baseline approach assumes that the force feedback, which is part of the state, implicitly provides all characteristic information relevant to the task. To improve adaptation, we propose to provide object characteristic information, such as Young's modulus, as part of the state. Such characteristic information are estimated from the force feedback during a predefined cut using a separate Neural Network (NN) trained by supervised learning.

III. Results and Conclusion

We conducted two experiments to test if explicitly inputting characteristic information improves the learning of adaptive cutting. For both experiments, we obtained 12 demonstrations of cutting an apple-shaped object by varying Young's modulus from 5.88×10^4 to $2.75 \times 10^6 Nm^{-2}$, using the differential cutting simulator DiSECt [3]. The collected demonstration data consists of state-action pairs. The state includes the end-effector position, velocity, and the contact force between the knife and the object. The contact force is the sum of the forces on all cutting springs inserted between the cutting surfaces computed by the simulator. The action of the robot is the end-effector velocity.

In the first experiment, the simulator generated demonstration data by minimising the contact force during cutting. The results showed that the normalised Mean Squared Error of the predicted force trajectories, which vary greatly in the demonstration data due to the different hardness of the objects, significantly reduced by 40% $(p < 0.05)$ compared to the baseline approach.

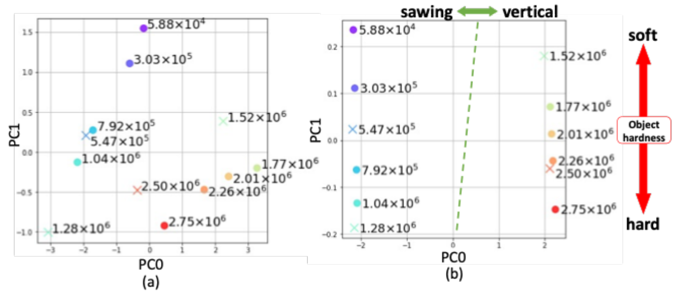


Fig. 1: PCA on the latent space during the task execution (a) without characteristic information and (b) with categorical characteristic information. The circle and cross marks show the training and validation data respectively. The values are the Young's modulus.

In the second experiment, we obtained demonstration data by performing a predefined vertical cut on the 6 objects with lower Young's modulus, and a sawing motion on the remaining ones. Here, instead of providing the continuous Young modulus values, a label of the 2 classes, soft or hard, was used. As shown in the results of applying Principal Component Analysis (PCA) to the NN latent space, providing the categorical characteristic enabled the NN to capture the correlation between the object characteristics (*i.e.,* hardness) and the different motions (*i.e.,* vertical or sawing cut) (see Fig. 1).

Our results indicate that explicitly providing characteristic information, estimated through force interaction, can bring additional information to adapt the learnt force trajectories which vary greatly depending on the characteristics of the manipulated objects. However, the improvement in the adaptation of the motion using our method from the baseline was not significant. Therefore, the most effective way to incorporate such characteristic information into NN-based Learning from Demonstration to achieve adaptive behaviour will be further investigated in future work.

References

[1] N. Saito, T. Ogata, S. Funabashi, H. Mori, S. Sugano, IEEE Robotics and Automation Letters 2021, 6, 2517. 2021. doi: 10.1109/LRA.2021.3062004.

[2] J. Nielsen, A. S. Sørensen, T. S. Christensen, T. R. Savarimuthu, T. Kulvicius, ICRA 2017 Workshop on Advances and challenges on the development, testing and assessment of assistive and rehabilitation robots: Experiences from engineering and human science research 2017, 1, 40.

[3] E. Heiden, M. Macklin, Y. Narang, D. Fox, A. Garg, F. Ramos, Robotics: Science and Systems XVII 2021, DOI 10.15607/rss.2021.xvii.067.

[1] Karlsruhe Institute of Technology

[2] Max Planck Institute for Intelligent Systems

[3] Chemnitz University of Technology

[4] Alpen-Adria-Universität Klagenfurt

[5] Carnegie Mellon University

[6] Queen Mary University of London

[7] Karlsruhe University of Applied Sciences

Karlsruhe Institute of Technology
Intelligent Process Automation and Robotics Lab (IPR)
Engler-Bunte-Ring 8
76131 Karlsruhe, Germany
www.ipr.kit.edu

Max Planck Institute for Intelligent Systems (MPI-IS)
Heisenbergstr. 3, 70569 Stuttgart
Stuttgart, Germany
www.is.mpg.de

Technische Universität Chemnitz
Straße der Nationen 62
09111 Chemnitz
https://www.tu-chemnitz.de

Alpen-Adria-Universität Klagenfurt
Universitätsstraße 65-67
9020 Klagenfurt am Wörthersee, Austria
www.aau.at

Carnegie Mellon University (CMU)
5000 Forbes Ave
Pittsburgh, PA 15213 USA
www.cmu.edu

Queen Mary University of London (QMUL)
Mile End Road
London, United Kingdom
www.qmul.ac.uk

Karlsruhe University of Applied Sciences
Moltkestr. 30
76133 Karlsruhe, Germany
www.hs-karlsruhe.de

Impressum

Karlsruher Institut für Technologie (KIT)
www.kit.edu