Rotational Slippage Prediction from Segmentation of Tactile Images

Julio Castaño-Amorós[1] and Pablo Gil[2]

***Abstract*— Adding tactile sensors to a robotic system is becoming a common practice to achieve more complex manipulation skills than those robotics systems that only use external cameras to manipulate objects. The key of tactile sensors is that they provide extra information about the physical properties of the grasping. In this paper, we implemented a system to predict and quantify the rotational slippage of objects in hand using the vision-based tactile sensor known as Digit. Our system comprises a neural network that obtains the segmented contact region (object-sensor), to later calculate the slippage rotation angle from this region using a thinning algorithm. Besides, we created our own tactile segmentation dataset, which is the first one in the literature as far as we are concerned, to train and evaluate our neural network, obtaining results of 95% and 91% in Dice and IoU metrics. In real-scenario experiments, our system is able to predict rotational slippage with a maximum mean rotational error of 3 degrees with previously unseen objects. Thus, our system can be used to prevent an object from falling due to its slippage.**

I. Introduction and Related Work

Traditionally, the methods to carry out robotic manipulation tasks used 2D or 3D vision sensors [1], which only take into account the geometric properties of the objects to perform the grasping. In contrast, with tactile sensors, it is possible to measure and react to physical properties (mass distribution, center of gravity or friction) in order to achieve a stable grasping [2].

In the last twenty years, several tactile sensors have been designed using different hardware technologies [3], although the last trend of tactile sensors lies in optical tactile sensors [4]. In this manuscript, we present an algorithm to estimate the rotation angle of an object which is being manipulated when slippage occurs. This method is based on segmentation neural networks to obtain the contact region (object-sensor) and traditional computer vision techniques to calculate the rotation angle and is applied to the vision-based tactile sensor Digit [5] which does not contain visual markers to keep low its cost.

Estimating the contact region between the robot's fingertips and the grasped object has been attempted to be solved in different ways. For example, by subtracting contact and no-contact tactile images [6], detecting and grouping visual markers [7], throughout 3D reconstruction and photometric algorithms [8], or using neural networks [9], [10]. In contrast, although our work is inspired by these previous articles, the main differences lie in the fact that we use the Digit sensors, without markers [7], which do not produce depth information [8], and state-of-the-art segmentation neural networks, which are more robust than subtracting operations [6] and vanilla CNN [9], and its training is more stable compared with GAN's training [10].

Slippage is a common physical event that occurs during object manipulation, that has been tried to solve for several years employing different approaches. For example, detecting binary slippage events with traditional image preprocessing techniques [11], combining convolutional and recurrent neural networks to classify slip in clockwise and counterclockwise rotation [12] or estimating the slip rotation angle using vision-based tactile sensors with markers [13] or force/torque sensors [14]. In this paper, we have inspired our work in these methods that characterize and quantify the rotational slip.

II. Method

We propose a two-stage method for touch region segmentation and rotational slippage prediction. The first stage of our method is based on a segmentation neural network applied to vision-based tactile sensing, which we called Tactile Segmentation Neural Network (TSNN). In this work, our goal is only to segment the contact region, then we decided to use DeepLabV3+ [15] architecture for experimentation. DeepLabV3+ is well-known for using an encoder-decoder architecture to perform image segmentation, and for introducing a new layer in its architecture, which is a combination of atrous or dilated and depth-wise separable convolutions. This combination leads to a reduction of computational complexity while maintaining similar or even better performance than previous versions. As the encoder, the authors used a modified version of the architecture Xception, called Aligned Xception, which replaces all the max pooling layers by the depth-wise separable convolutions to perform the feature extraction procedure. The decoder, in contrast, is a simpler part of the architecture, which only comprises convolution, concatenation, and upsampling layers to transform the intermediate features into the output.

The second stage of our method estimates the angle of rotation of the segmented region of contact using a traditional computer vision thinning algorithm (Skeleton method) [16] that blackens points in the binary contact region using an 8-square neighborhood and different connectivity conditions. Other approaches, based on different neural networks such

*This research was funded by the Valencian Regional Government through the PROMETEO/2021/075 project and by the University of Alicante through the grant UAFPU21-26

[1]Julio Castaño-Amorós is with University Institute for Engineering Research, Miguel Hernández University, Elche and with AUROVA Lab, Computer Science Research Institute, University of Alicante, Alicante, Spain julio.ca@ua.es

[2]Pablo Gil is with AUROVA Lab, Computer Science Research Institute and with Department of Physics, Systems Engineering, and Signal Theory, University of Alicante, Alicante, Spain pablo.gil@ua.es

as Unet++ [17] and PSPNet [18] or different algorithms to estimate the angle such as PCA or ellipse fitting, were tested. The complete system is shown in Fig 1.

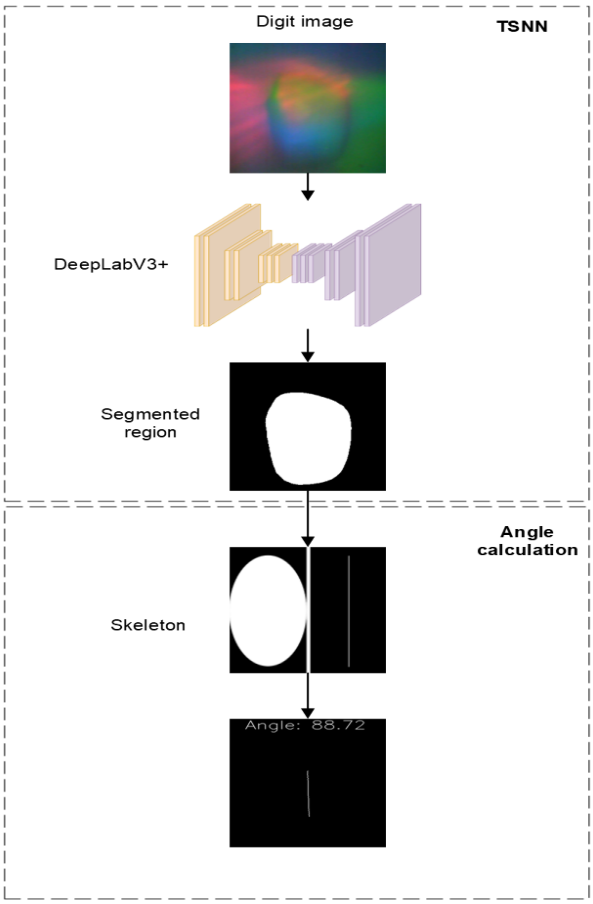


Fig. 1: Diagram of our system combining both stages

III. Experimentation and Results

We have generated our own dataset as we have not found any dataset related to tactile segmentation in order to be used as the base of our experimentation. Our tactile segmentation dataset comprises 3675 tactile images with their respective labelled contact regions. We have used 16 objects from YCB dataset to record it, from which we have captured between 200 and 250 tactile images per object. The objects contain different textures, rigidity, weight, geometries, etc.

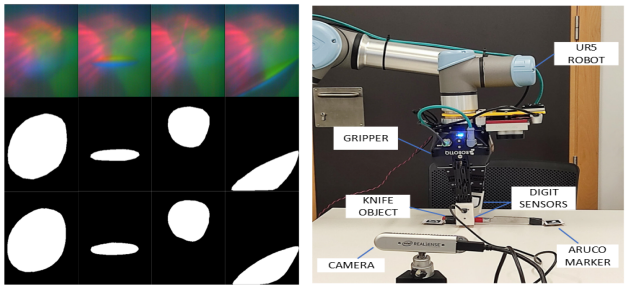
To train the TSNN we use the Dice and IoU metrics, an NVIDIA A100 Tensor Core GPU with 40 GB of RAM, and the following optimal hyperparameters: a batch size of 32, a learning rate of 1e-4, the Adam optimizer, the Focal loss, and 30 training epochs.

Table I shows the results obtained by DeepLabV3+ TSNN in the testing experiment. DeepLabV3+ is able to segment tactile images with high accuracy and in real-time execution. Besides, this TSNN is 3 ms faster than other segmentation neural networks (Unet++ and PSPNet) while maintaining the same performance, thus, achieving a better trade-off between segmentation accuracy and prediction time.

TABLE I: DeepLabV3+ TSNN results in terms of Dice, IoU and inference time metrics, and using the backbone ResNet18

	Dice	IoU	Time(s)
DeepLabV3+	0.956 ± 0.013	0.914 ± 0.023	0.006 ± 0.002
PSPNet	0.951 ± 0.014	0.907 ± 0.025	0.006 ± 0.002

Figure 2a shows different examples of contact region segmentation carried out by DeepLabV3+ TSNN, and Fig. 2b shows our robotic manipulation setup with a UR5 robot, two DIGIT sensors, a ROBOTIQ gripper, the object to grasp with the aruco markers attached and an Intel RealSense camera to calculate the ground truth angle.



(a) (b)

Fig. 2: a) Examples of rotation angle calculation for slipping during lift task: DIGIT image (first row), ground truth (second row), prediction (third row), b) Robotic manipulation setup with different objects

The task consists of grasping and lift an object while the tactile segmentation and rotational slippage angle are estimated. The predicted angle is calculated as the difference between the current and the initial angle obtained in the Skeleton method described earlier, while the ground truth angle is calculated using two aruco markers as visual references. Our system was evaluated with seven unseen objects (1 to 7 in Fig. 3) and two seen objects from our tactile segmentation dataset (8 and 9 in Fig. 3).

The experimentation comprises 45 graspings and lifts in total (five per object) while calculating the rotational error in degrees. Figure 3 shows the mean rotational error of the 5 graspings and lifts for each object. Note that object 6 and 8 causes more error and deviation because object 6 weight's is higher compared with the rest of the objects, and object 8 contains higher curvature on its surface that causes more saturation in the sensor. Our system is able to predict rotational slippage with an overall mean rotational error of **1.854º** ± **0.988º**, that is to say, a maximum mean error of 3 degrees in the worst case. Figure 4 shows some examples of the prediction of rotational slippage with four aforementioned objects.

IV. Conclusions

In this paper, we propose a model-based system to predict rotational slippage during the grasping and lift of an object, achieving a mean error value of **1.854º** ± **0.988º**, compared

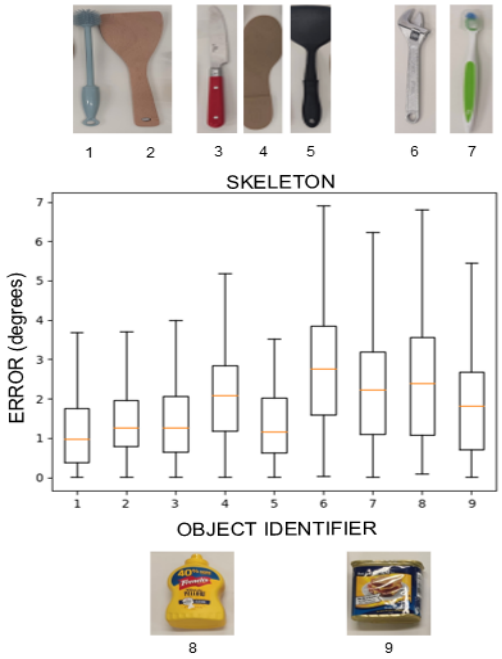
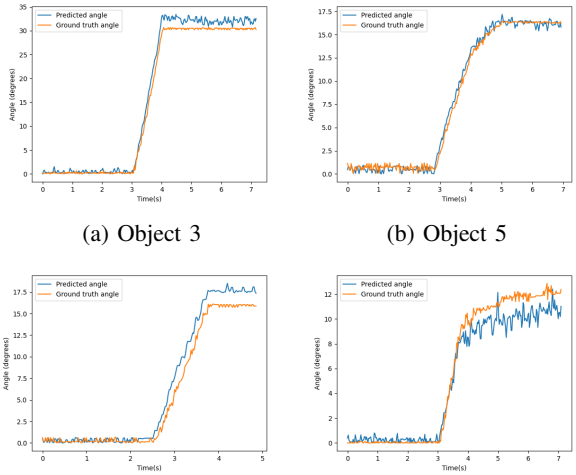


Fig. 3: Rotation errors for each object using the Skeleton method



(a) Object 3 (b) Object 5

(c) Object 4 (d) Object 1

Fig. 4: Examples of rotation angle calculation for slipping during lift task with different objects

with the error of **3.96º** ± **UNK** from [13], and the error of **4.39º** ± **0.18º** from [14]. Although we could not carry out an experimental comparison because we do not have their sensors available, some objects were used both in this work and in theirs. Our system also has some limitations regarding the shape of the contact region. If this shape is similar to a circle, it becomes impossible to calculate its rotation movement. In that case, we propose to grasp the object by surfaces with small curvature.

REFERENCES

[1] G. Du, K. Wang, S. Lian, and K. Zhao, "Vision-based robotic grasping from object localization, object pose estimation to grasp estimation for parallel grippers: a review". Artificial Intelligence Review, vol. 54, pp. 1677–1734, 2021, doi: 10.1007/s10462-020-09888-5

[2] S. Luo, J. Bimbo, R. Dahiya and H. Liu, "Robotic tactile perception of object properties: A review". Mechatronics, vol. 48, pp. 54-67, 2017, doi: 10.1016/j.mechatronics.2017.11.002

[3] C. Chi, X. Sun, N. Xue, T. Li, and C. Liu. "Recent Progress in Technologies for Tactile Sensors", Sensors, vol. 18, no. 4, page 948, doi: 10.3390/s18040948

[4] S. Zhang et al., "Hardware Technology of Vision-Based Tactile Sensor: A Review," in IEEE Sensors Journal, vol. 22, no. 22, pp. 21410-21427, 15 Nov.15, 2022, doi: 10.1109/JSEN.2022.3210210

[5] M. Lambeta et al., "DIGIT: A Novel Design for a Low-Cost Compact High-Resolution Tactile Sensor With Application to In-Hand Manipulation," in IEEE Robotics and Automation Letters, vol. 5, no. 3, pp. 3838-3845, July 2020, doi: 10.1109/LRA.2020.2977257

[6] M. Lambeta, H. Xu, J. Xu, P. W. Chou, S. Wang, T. Darrell, and R. Calandra, "PyTouch: A Machine Learning Library for Touch Processing," in 2021 IEEE International Conference on Robotics and Automation (ICRA), Xi'an, China, 2021, pp. 13208-13214, doi: 10.1109/ICRA48506.2021.9561084

[7] Y. Ito, Y. Kim, G. Obinata, "Contact Region Estimation Based on a Vision-Based Tactile Sensor Using a Deformable Touchpad," Sensors, vol.14, no. 4, pp. 5805-5822, 2014, doi: 10.3390/s140405805

[8] S. Wang, Y. She, B. Romero and E. Adelson, "GelSight Wedge: Measuring High-Resolution 3D Contact Geometry with a Compact Robot Finger," 2021 IEEE International Conference on Robotics and Automation (ICRA), Xi'an, China, 2021, pp. 6468-6475, doi: 10.1109/ICRA48506.2021.9560783

[9] M. Bauza, O. Canal and A. Rodriguez, "Tactile Mapping and Localization from High-Resolution Tactile Imprints," 2019 International Conference on Robotics and Automation (ICRA), Montreal, QC, Canada, 2019, pp. 3811-3817, doi: 10.1109/ICRA.2019.8794298

[10] Y. Lin, J. Lloyd, A. Church and N. F. Lepora, "Tactile Gym 2.0: Sim-to-Real Deep Reinforcement Learning for Comparing Low-Cost High-Resolution Robot Touch," in IEEE Robotics and Automation Letters, vol. 7, no. 4, pp. 10754-10761, Oct. 2022, doi: 10.1109/LRA.2022.3195195

[11] J. Castaño-Amorós, I. de L. Páez-Ubieta, P. Gil, S. Puente, "Visual-tactile manipulation to collect household waste in outdoor". Revista Iberoamericana de Automática e Informática Industrial, 00, pp. 1-12, 2022, doi: 10.4995/riai.2022.18534

[12] B.S. Zapata-Impata, P. Gil and F. Torres, "Learning Spatio Temporal Tactile Features with a ConvLSTM for the Direction Of Slip Detection" in Sensors, vol. 19, no. 523, 2019, doi: 10.3390/s19030523

[13] R. Kolamuri, Z. Si, Y. Zhang, A. Agarwal, and W. Yuan, "Improving Grasp Stability with Rotation Measurement from Tactile Sensing," in 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Prague, Czech Republic, 2021, pp. 6809-6816, doi: 10.1109/IROS51168.2021.9636488

[14] J. Toskov, R. Newbury, M. Mukadam, D. Kulić, and A. Cosgun, "In-Hand Gravitational Pivoting Using Tactile Sensing", arXiv preprint arXiv:2210.05068, 2002

[15] L. C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 801-818, doi: 10.1007/978-3-030-01234-2_49

[16] Z., GUO and R. W., HALL, "Parallel thinning with two-subiteration algorithms," Communications of the ACM, 1989, vol. 32, no. 3, pp. 359-373, doi: 10.1145/62065.62074

[17] Z. Zhou, M.M. Rahman Siddiquee, N. Tajbakhsh, J. Liang. "UNet++: A Nested U-Net Architecture for Medical Image Segmentation," in Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. DLMIA ML-CDS 2018 2018. Lecture Notes in Computer Science(), vol 11045. Springer, Cham, doi: 10.1007/978-3-030-00889-5_1

[18] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid Scene Parsing Network," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 6230-6239, doi: 10.1109/CVPR.2017.660