vision2tactile: Feeling Touch by Sight

1st Brayan S. Zapata-Impata
AUROVA Lab
University of Alicante
Alicante, Spain
brayan.impata@ua.es

2nd Pablo Gil
AUROVA Lab
University of Alicante
Alicante, Spain
pablo.gil@ua.es

3rd Fernando Torres
AUROVA Lab
University of Alicante
Alicante, Spain
fernando.torres@ua.es

***Abstract*—Latest trends in robotic grasping combine vision and touch for improving the performance of systems at tasks like stability prediction. However, tactile data are only available during the grasp, limiting the set of scenarios in which multi-modal solutions can be applied. Could we obtain it prior to grasping? We explore the use of visual perception as a stimulus for generating tactile data so the robotic system can "feel" the response of the tactile perception just by looking at the object.**

I. Introduction

Whenever we humans grasp objects, both visual and tactile perception play a paramount role: vision provides us information of the object like its geometry and touch lets us discover its stiffness, among other properties. As avid learners, we can estimate physical properties of novel objects just by looking at them. It is argued [3] that our brain builds statistical generative models, which capture the visual attributes of textures or materials, so we can predict how a surface would feel if we touch it. In this fashion, we explore how a robot could learn to model its tactile sense using visual perception, so that it can estimate the tactile response of its sensors. More precisely, we aim to generate the tactile data that would be registered if the robot grasps an object, given a 3D point cloud of it.

Previous works have shown the importance of tactile perception in robotic grasping. Calandra *et al.* [1] trained a CNN with both tactile and visual images in order to predict grasp success. Regrasp has been recently approached using simulated tactile data as well [4], mapping generated tactile images into actions of the robotic gripper. However, their systems needed to grasp the object in order to obtain a tactile image. Recently, Lee *et al.* [5] trained a cross-modal generative model that produced tactile images from real visual images of textures and vice versa. Our work is similar to this one, but with significant differences: 1) we use tactile sensors that record signals instead of tactile images so our data are of different kind, 2) our input is also of a different type (3D point cloud), and 3) we have recorded a dataset of real grasps with visual and tactile data.

II. Robotic System

In our setting, we use the BioTac SP tactile sensor developed by SynTouch. It holds 24 electrodes distributed throughout its internal core, which record signals from 4 emitters and measure the impedance in the fluid located between them and the elastic skin of the sensor. As a result, the greater the pressure, the lower the voltage readings of the electrodes. In addition, the sensor provides a global pressure measurement using a sensor in its base. We work with two BioTac SP sensors installed on a Shadow Dexterous Hand (middle finger and thumb). Besides, we use the Intel RealSense D415 depth camera, that records dense 3D point clouds, fixed in the world in an eye-to-hand configuration.

Work funded by the Spanish Government and the FEDER Funds (BES-2016-078290, RTI2018-094279-B-100) as well as by the European Commission (COMMANDIA SOE2/P1/F0638), action supported by Interreg-V Sudoe.

III. Touch Modelling

We propose to model the BioTac SP tactile sensors using visual perception and deep neural networks. In detail, a 3D point cloud of the object to be grasped is fed to a network, which outputs the tactile readings that would be registered if the grasp was performed. We identify three key questions: A) how should we represent the object? B) what should be the output of the network? and C) what should be the architecture of such network that will model the sensor's behaviour?

A. Visual Representation

For representing the object, we propose 3D point clouds. This structure represents the geometry of the object, which could be useful for modelling the tactile response. Besides, we segment the object from the background so the cloud $\mathbb{C} = \{p_1, p_2, ..., p_n\}$ only holds points that belong to the object. Moreover, only their 3D coordinates $p_i = (x, y, z)$ are used.

B. Tactile Responses

Two levels of difficulty are identified for this task: in the simpler version of the task, the system has to learn to generate the global pressure value, called *DC pressure* or PDC. Therefore, it must regress two values: PDC_{mf} and PDC_{th}, one for each sensor. In the more complex version, it has to learn to generate the readings for each of the 24 electrodes. As a result, it must learn to regress 24 values for each sensor: $E_{mf} = \{e_1, e_2, ..., e_{24}\}$ and $E_{th} = \{e_1, e_2, ..., e_{24}\}$.

C. Network Architecture

We propose the use of a network based on PointNet [7] for calculating deep features from the 3D point clouds and use them for regressing tactile readings. Since point clouds can have different sizes, we downsample them to 500 points and normalise their coordinates to the unit sphere with centre at

the point cloud's centroid. As for tactile responses, they are scaled from their discrete values to the continuous range $[0, 1]$. These normalisation were necessary for convergence reasons.

IV. Data Collection

In order to train such a network, we needed to record a dataset of real grasps. Grasps were executed with the Shadow Hand mounted on a Mitsubishi PA10 robotic arm. We used GeoGrasp [8] for computing grasping points on the 3D point clouds. From every grasp, we saved the 3D point cloud of the object and the tactile readings experienced at the moment of contact, so a sample is a tuple $S = \langle \mathbb{C}, PDC_{mf}, PDC_{th}, E_{mf}, E_{th} \rangle$. The robotic fingers were closed so they would contact the object on the computed grasping points applying sufficient force but without exceeding the torque limits of the joints. In consequence, we have recorded a set of grasps with different tactile responses. Grasps were executed on a set of objects from the YCB object set [2]. We used two cylinder-like objects (can of crisps and can of coffee) and two box-like objects (snacks box and sugar box), executing 50 grasps per object[1].

V. Experiments

As can be seen in Figure 1, the range of values of the sensors are different: the sensor on the middle finger does not exceed values above 2000 – these are custom units used by the BioTac SP sensor – while the thumb's sensor is always over that. Due to construction reasons, the sensors behave differently under the same conditions. Hence, we trained one network for each sensor, though sharing architecture and hyper-parameters. Networks were trained optimising the Root Mean Squared Error (RMSE) and 5-fold cross-validation was carried out.

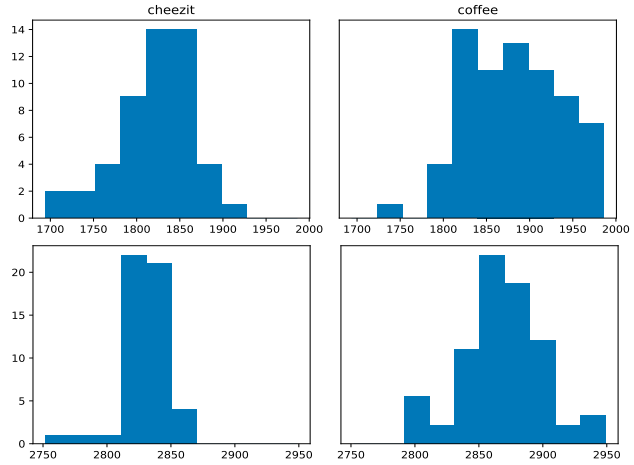

Fig. 1. Distribution of (top) PDC_{mf} and (bottom) PDC_{th} for two objects.

A. Regressing Global Pressure

Training on the cylinder-like objects yielded an average RMSE of 0.076 and 0.067 for PDC_{mf} and PDC_{th} respectively. Scaling those errors back to the sensors' range, they equal 153 and 208 units. Similarly, training on box-like objects, the errors are 0.081 (164 units) and 0.089 (278 units) for each sensor. This shows that it is possible to learn to regress PDC using PointNet. However, the error is still large, just as large as the range of values for each of the sensors (see ranges in Figure 1). Finally, training with one type of object and testing on the other kept the mean error at 165 and 236.

B. Regressing Electrodes Readings

In this case, training on cylinder-like objects yielded an average RMSE of 0.060 (231 units) and 0.061 (216 units) for E_{mf} and E_{th}. Training on box-like objects, we obtained errors equal to 0.055 (212 units) and 0.066 (232 units). As expected, the errors are higher since the network had to learn to regress more values. This tendency was more evident training the system with one type of object and testing on the other: the average errors rose to 310 and 317 points for each sensor.

VI. Discussion and Open Opportunities

Visual perception is currently being simulated for learning manipulation policies that can be transferred to real systems [6]. With this work, we aim to give a rich source of information to those agents because tactile perception should be paramount for learning manipulation skills. 3D point clouds could be generated from synthetic depth images and then fed to our system for regressing a tactile response. Therefore, it would act as a simulation of the sensor created with real data.

In future lines, we want to test further the performance of our proposal on objects with more diverse geometries and degrees of stiffness. Moreover, we would like to test its behaviour with synthetic 3D point clouds. In addition, we want to compare it against a GAN-based approach, given their excellent performance at similar tasks [5] and the inspiration found in the way our brain builds visual models [3].

[1]Data available at: https://github.com/yayaneath/vision2tactile

References

[1] Roberto Calandra, Andrew Owens, Manu Upadhyaya, Wenzhen Yuan, Justin Lin, Edward H. Adelson, and Sergey Levine. The Feeling of Success: Does Touch Sensing Help Predict Grasp Outcomes? In *1st Annual Conference on Robot Learning*, volume 78, pages 314–323, 2017.

[2] Berk Calli, Arjun Singh, James Bruce, Aaron Walsman, Kurt Konolige, Siddhartha Srinivasa, Pieter Abbeel, and Aaron M Dollar. Yale-CMU-Berkeley dataset for robotic manipulation research. *The International Journal of Robotics Research*, 36(3):261–268, 2017.

[3] Roland W. Fleming. Visual perception of materials and their properties. *Vision Research*, 94:62–75, 2014.

[4] Francois R. Hogan, Maria Bauza, Oleguer Canal, Elliott Donlon, and Alberto Rodriguez. Tactile Regrasp: Grasp Adjustments via Simulated Tactile Transformations. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2963–2970. IEEE, oct 2018.

[5] Jet-Tsyn Lee, Danushka Bollegala, and Shan Luo. "Touching to See" and "Seeing to Feel": Robotic Cross-modal SensoryData Generation for Visual-Tactile Perception. feb 2019.

[6] OpenAI, Marcin Andrychowicz, Bowen Baker, Maciek Chociej, Rafal Jozefowicz, Bob McGrew, Jakub Pachocki, Arthur Petron, Matthias Plappert, Glenn Powell, Alex Ray, Jonas Schneider, Szymon Sidor, Josh Tobin, Peter Welinder, Lilian Weng, and Wojciech Zaremba. Learning Dexterous In-Hand Manipulation. 2018.

[7] Charles R. Qi, Hao Su, Kaichun Mo, and Leonidas J. Guibas. PointNet: Deep learning on point sets for 3D classification and segmentation. *30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017-Janua:77–85, 2017.

[8] Brayan S. Zapata-Impata, Pablo Gil, Jorge Pomares, and Fernando Torres. Fast geometry-based computation of grasping points on three-dimensional point clouds. *Int. J. of Advanced Robotic Systems*, 16(1), jan 2019.