Accurate estimation of the 3D contact force distribution with an optical tactile sensor – Live demonstration

Carmelo Sferrazza and Raffaello D'Andrea

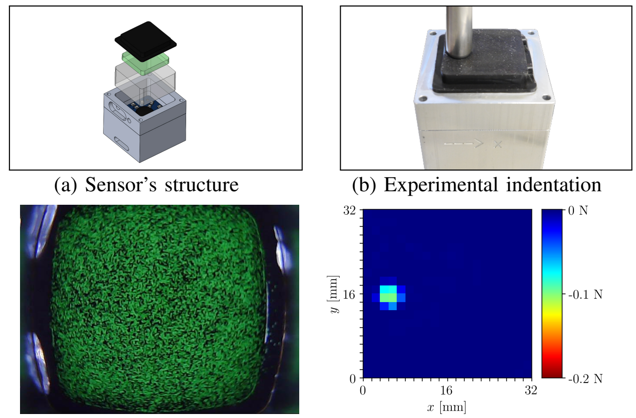
***Abstract*— This paper describes a tactile sensing pipeline that aims to reconstruct the contact force distribution applied to a soft surface. The sensing principle is based on a camera that tracks the motion of plastic particles within a deformable material. Optical flow features are extracted from the images and mapped to contact force distributions via a neural network architecture. Two strategies for collecting the training data are described, one of which can entirely be carried out in simulation, while retaining most of the accuracy, as shown when the architecture is evaluated on real-world sensors. The simplicity of the hardware components employed and the efforts towards real-time software implementation provide a platform that is reliably available for live demonstrations.**

I. INTRODUCTION

In recent years, vision-based tactile sensors have increasingly been employed for robotics applications, due to their low cost, ease of manufacture and high resolution. While promising results have been obtained for different tasks, the development of a pipeline that accurately reconstructs the contact pressure and shear forces applied to the sensing surface for general cases has up till now not been achieved. In fact, while several techniques have been developed for the estimation of total forces, contact shapes and locations [1], these are generally limited to specific cases, e.g., indentations with single objects. In contrast, a comprehensive estimation of the contact properties may be provided by the 3D contact force distribution. The quantities mentioned above (total forces, etc.) can be extracted from the force distribution, which also provides a way of handling a variety of contact conditions (e.g., contact with multiple objects). The versatility of the contact force distribution makes it a very appealing feedback quantity for generic robotic manipulation tasks.

While first-order accuracy can be obtained with model-based techniques [2] by assuming linear elasticity of the soft materials, it is generally not feasible to account for the hyperelastic behavior of rubber in real-time. In order to overcome this limitation, data-driven methods can be trained on arbitrarily accurate data, while retaining short inference times. However, since there are no readily available sensors that can measure the force distribution applied to soft materials without altering the contact surface, the collection of data needed for training is not straightforward.

This paper is based on a tactile sensing principle originally presented in [3], which uses a camera to track the motion of green particles (see Fig. 1, (a) and (c)) randomly spread within a soft, deformable material. The information provided by the images about the deformation of the sensing surface can be mapped to the contact force distribution via a data-driven technique. Note that, as opposed to the tracking of sparse markers, this information is retrieved at each pixel of the image, since the particles are densely spread within the deformable material. Here, two methods to collect the training data are compared: in the first method, optical flow features are obtained from experimental indentations and matched to highly accurate ground truth force distributions, obtained through performing the same indentations in a simulation environment based on the finite element method (FEM); in the second method, both the optical flow features and the ground truth labels are obtained in simulation, where a camera projection model is additionally employed. The complete explanations of the two methods can be found in [4] and [5], respectively. The two resulting datasets are separately used to train an artificial neural network, which yields real-world, accurate predictions at 50 Hz on the CPU of a laptop computer. The evaluation experiments show that the need for real-world training data can be greatly reduced. Furthermore, the learning architecture retains sensible predictions on data different from those seen during training.



(a) Sensor's structure (b) Experimental indentation (c) Internal camera image (d) Normal force distribution

Fig. 1: The images above illustrate the structure of the sensor employed and the different steps of the sensing pipeline.

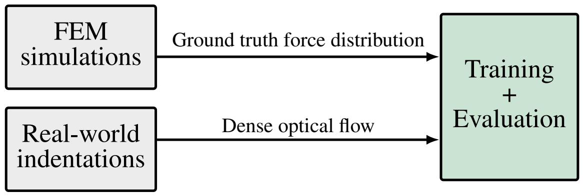
II. METHOD

In order to train a data-driven architecture that aims to reconstruct the force distribution applied on the soft surface, the task is formulated as a supervised learning problem. In the following subsections, two different methods to generate training and evaluation datasets for such a task are described.

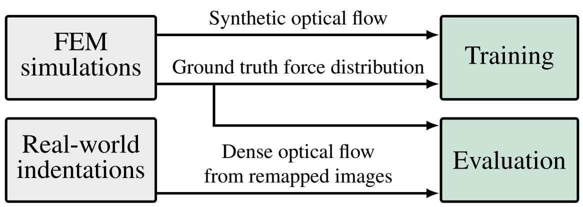
A. *Mixed-source dataset*

Thousands of vertical indentations (see Fig. 1(b)) are performed at various depths over different locations of the

The authors are members of the Institute for Dynamic Systems and Control, ETH Zurich, Switzerland. Email correspondence to Carmelo Sferrazza `csferrazza@ethz.ch`



(a) Mixed-source data collection



(b) Sim-to-real data collection

Fig. 2: The diagrams outline the two methods for data collection.

sensing surface using a spherical indenter attached to an automatic milling machine. During these experiments, the images from the internal camera are recorded. These images are processed offline by extracting the dense optical flow (based on [6]) with respect to an image frame taken at rest. Note that, prior to the optical flow extraction, the images are converted to gray-scale, as required by the algorithm employed. The optical flow is then further subsampled, performing an average pooling over a 40×40 grid for each component.

In order to overcome the lack of commercial sensors measuring the contact force distribution, ground truth labels are assigned to the optical flow features corresponding to each image by repeating the same indentations in an FEM simulation environment. From the simulations, three-channel 20×20 grids of labels are extracted, representing the discretized contact force distributions. The simulations are based on hyperelastic models of the materials employed, obtained via state-of-the-art characterization techniques. These models exhibit high accuracy, as detailed in [4]. For the sake of simplicity, time-dependent behaviors of the material, e.g., hysteresis and wear, were not considered in this work. However, complex models accounting for these behaviors could be employed in a similar manner to generate ground truth labels.

The resulting dataset, comprising optical flow features and corresponding ground truth labels, is randomly split into training and evaluation datasets. A diagram describing the procedure is shown in Fig. 2.

B. Sim-to-real dataset

By simulating the camera projection, the collection of real-world data can be avoided, replacing the optical flow features extracted from the real images with features generated in simulation. The procedure is explained in detail in [5], and is based on the projection of the displacement field extracted from the FEM simulations onto the image plane, assuming an ideal pinhole camera model. Ground truth labels are assigned to these features as described in the previous subsection.

An evaluation dataset is collected by performing real-world indentations on a tactile sensor. The images extracted are remapped as if they were taken with the pinhole camera assumed in simulation. For this purpose, standard camera calibration techniques are employed before finalizing the sensor production, and a single real-world indentation is employed to refine the remapping procedure. The optical flow is then extracted from the remapped images, as outlined in Fig. 2.

III. RESULTS

The two datasets described in Section II are separately used to train a neural network architecture, which is inspired by u-net [7] and is described in [5]. The results on the respective evaluation datasets (with vertical forces up to 1.7 N) are shown in Table I for the horizontal (F_x and F_y) and vertical (F_z) components of the force distribution. Both the models trained exhibit high accuracy in estimating the contact force distribution and the total force applied. While the mixed-source dataset performs slightly better, the sim-to-real model only required a single real-world indentation and can be transfered across multiple instances of the sensors produced, provided that the camera calibration has been performed.

A sample prediction of the network trained on the mixed-source dataset is shown in Fig. 1(d). The experimental evaluation of the sim-to-real dataset is available in the linked video[1], which shows how the convolutional structure of the network yields sensible generalization capabilities when applied to cases with multiple points of contact and indenters of different shapes.

The sensing pipeline reconstructs the contact force distribution in real-time at 50 Hz. The system is therefore readily available for live demonstrations, streaming the camera input, the optical flow features and the corresponding reconstruction of the force distribution to a screen.

Metric	F_x	F_y	F_z
RMSE - MS	0.001 N	0.001 N	0.001 N
RMSE total force - MS	0.015 N	0.017 N	0.033 N
RMSE - S2R	0.001 N	0.001 N	0.004 N
RMSET total force - S2R	0.032 N	0.041 N	0.131 N

TABLE I: The table summarizes the resulting errors for the models trained on the mixed-source (MS) and sim-to-real (S2R) datasets.

REFERENCES

[1] K. Shimonomura, "Tactile Image Sensors Employing Camera: A Review," *Sensors*, 2019.
[2] D. Ma, E. Donlon, S. Dong and A. Rodriguez, "Dense Tactile Force Estimation using GelSlim and inverse FEM," *ICRA*, 2019.
[3] C. Sferrazza and R. D'Andrea, "Design, Motivation and Evaluation of a Full-Resolution Optical Tactile Sensor," *Sensors,* 2019.
[4] C. Sferrazza, A. Wahlsten, C. Trueeb and R. D'Andrea, "Ground Truth Force Distribution for Learning-Based Tactile Sensing: A Finite Element Approach," *IEEE Access*, 2019.
[5] C. Sferrazza, T. Bi and R. D'Andrea, "Learning the sense of touch in simulation: a sim-to-real strategy for vision-based tactile sensing," *arXiv:2003.02640*, 2020.
[6] T. Kroeger, R. Timofte, D. Dai, and L. Van Gool, "Fast optical flow using dense inverse search," *ECCV*, 2016.
[7] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," *MICCAI*, 2015.

[1]Link to the video [5]: `https://youtu.be/dDTga9PgWS0`