A data-driven approach to human-robot co-manipulation of soft materials

Giorgio Nicola, Enrico Villagrossi, Nicola Pedrocchi

***Abstract*—Human-robot co-manipulation of large but lightweight elements made by soft materials is a challenging operation that presents several relevant industrial applications. This paper proposes using a 3D camera to track the deformation of soft materials for human-robot co-manipulation. Thanks to a Convolutional Neural Network (CNN), the acquired depth image is processed to estimate the element deformation. The output of the CNN is the feedback for the robot controller to track a given set-point of deformation.**

I. Introduction

The human-robot co-manipulation of soft materials is becoming a relevant task from the industrial point of view, such as in aerospace, transport, maritime industries. Compared to the manipulation of rigid materials, it introduces new challenges in modeling, perception, grasping, and control [1]. In [2], the user guides the IMM (Industrial Mobile Manipulator) through gestures recorded by a camera and translated into robot control signals using a skeleton tracking algorithm and force feedback. Manipulating deformable materials in collaboration with humans or without (often called shape servoing) can be done with model-based [3]–[6] and model-free [7]–[10] approaches. In model-based approaches, a physics-based or black-box model describes the material's mechanical status (*e.g.* deformations, internal stress, etc.). Instead, model-free methods focus on developing handcrafted visual features to be converted directly into robot commands. This paper proposes to learn a model describing the displacement-deformation relation through a neural network, taking a depth map from a 3D camera as input. The model is used online to determine the displacement from a nominal configuration, and the displacement is fed to a Twist controller. This approach, compared to methods in the literature, has various advantages: 1) it is very straightforward compared to those described in the literature, such as those based on the deformation Jacobian matrix; 2) it does not require manually developing visual features that might not describe the desired problem fully.

II. Method

A. Problem Formulation

Soft materials like textiles can be approximated as membranes [11] characterized by the absence of flexural rigidity and cannot sustain compressive loads. Therefore, deformations can be caused only by displacements or by traction forces. Assuming neglectable traction forces, the material shape is univocally defined as the relative distance between the robot and the human grasping points. As shown in Figure 1, a nominal shape of the carbon fibre ply is defined as a human-robot reference displacement $(d_x^{rif}, d_y^{rif}, d_z^{rif})$ and the objective is to compute the necessary robot displacement $(\Delta d_x, \Delta d_y, \Delta d_z)$ to reach the nominal shape. We propose a data-driven black-box model composed of an ensemble of CNNs that, given as input a depth image from the robot point of view, computes the current human-robot displacement (d_x, d_y, d_z).

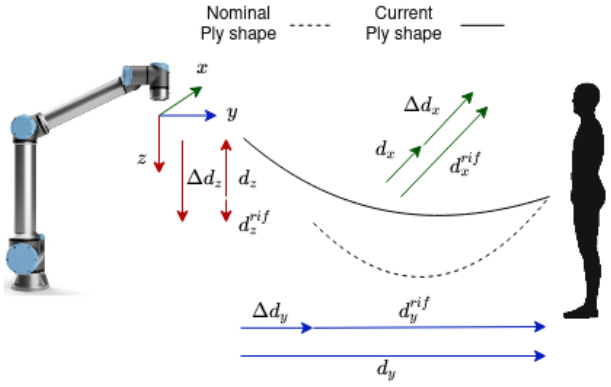

Fig. 1. Problem formulation.

B. Dataset Acquisition

The dataset to be acquired consists of multiple depth images of a carbon fiber ply deformed due to the human-robot relative displacement.

To increase the accuracy and repeatability of the measurements and the robustness of the trained model, the human is substituted with a frame that allows simulation of different human grasping positions.

The frame position is estimated via a pair of fiducial markers (Apriltags [12]), and the robot moves relatively to the frame in the various studied directions. In each robot pose, multiple RGB-D images are acquired to lower the noisy camera output. The corresponding label is the distance in the three directions for each dataset entry.

C. Neural Network and Training

The model takes inspiration from VGG16 [13]. The net shares the same general architecture based on blocks of two convolutional layers interspersed by batch normalization and then a maxpool layer. After those blocks, fully-connected layers combined with dropout layers are implemented, and

Giorgio Nicola *et al.* are with the Institute of Intelligent Industrial Technologies and Systems for Advanced Manufacturing, National Research Council of Italy, Via A. Corti 12, 20133, Milan, Italy; *giorgio.nicola@stiima.cnr.it*

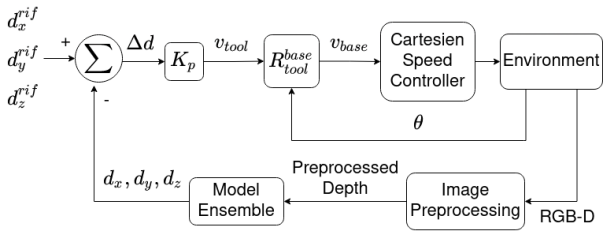

Fig. 2. Control scheme.

the net output is the Cartesian distance along with the three directions.

In detail, we trained three slightly different nets on different subsets of the dataset, given the fact that multiple depth images are taken for each relative robot-human position, The dataset is divided into two separate datasets for each net: training (80%) and test (20%) datasets. Such choice allows verification during the model's training to generalize over unseen relative robot-human positions.

We used Optuna [14] to optimize the hyperparameters of each net, implementing the K-fold cross-validation (K=5). The outputs from the various nets are combined by averaging. As data augmentation, pepper and Gaussian noises and random translations are applied.

D. Robot Control

The robot control scheme is described in Figure 2. After the image preprocessing, the model ensemble (*i.e.*, ensemble of CNNs) outputs the estimated human-robot distance, *i.e.*, the ply deformation, and a proportional controller converts the error in ply deformation, Δd, to a tool velocity in the tool frame. To avoid excessive tool velocities, v_{tool} is saturated to a maximum of 5 $[cm/s[$ in every direction. Finally, the tool velocity is converted from the tool frame to the robot base frame through the rotation matrix R_{tool}^{base} analytically computed from the robot joints angle θ. The control frequency is 7 $[Hz]$, higher than the average human reaction time.

III. Experiments

The Section reports the results of two different tests [1]. First, we analyze the step response to a ply deformation. Then, we analyze a manual guidance operation.

1) Step Response Analysis: the ply was attached to the frame used during the dataset acquisition. The robot starting position was displaced of a known value from the ply resting configuration set to $(d_x = 0, d_y = 0.6, d_z = 0)$.

Figures II-D and II-D show the results of two different trials. In both cases, the robot reaches the ply rest configuration successfully.

The bottom graphs in Figures II-D and II-D report the estimation error and its relation with the robot tool speed. The error is maximum at the beginning, decreasing accordingly with the robot tool speed. Indeed, as detailed in Section II-D, the vision system (camera and preprocessing) runs approximately at 30 $[Hz]$, while the controller receives the averaged last three frames. On the one hand, averaging the depth images reduces noise and, therefore, improves the stability of the estimation of the ply deformation. On the other hand, the average depth image becomes slightly blurred when the robot tool is high, and the estimation accuracy decreases.

Nevertheless, the inaccuracy of the estimation is still consistent with the actual ply deformation, *i.e.*, it never estimates a deformation in the opposite direction of the real one. Furthermore, the developed system can recover from the inaccurate estimation and converge to the desired ply deformation.

Figure 3 shows the estimation robustness when the deformation is beyond the limits of the dataset acquisition campaign. Indeed, the initial deformation in x and z directions are 0.18 and 0.23 $[m]$. Even though the estimation is somehow inaccurate in the z-direction, it is still consistent, and the system can converge to an accurate to the desired ply deformation.

2) Manual Guidance: finally, in Figure 4 we studied the case of manual guidance. The human was required to perform four movements of arbitrary lengths in the direction $x \rightarrow z \rightarrow y \rightarrow x$ highlighted between the green (start of the human movement) and the red (end of the human movement) dashed lines. The robot could follow human instructions in all cases, and the robot movements were smooth. Even in this scenario, the human could efficiently perform movements that would require the ply to deform beyond the limits in the dataset, confirming the robustness of the approach.

IV. Conclusions and Future Works

This paper proposes a Data-driven method for human-robot co-manipulation of flexible materials. The method implements black-box model, based on an ensemble of deep neural networks, that estimates current relative human-robot displacement from depth images. Subsequently, the displacement error to a reference displacement turns into Twist command. The paper also describes the methodology used to acquire the dataset, preprocess it, and train the ensemble model. The proposed method achieved an overall mean average error of 0.0215 $[m]$, and it requires a computation time, including preprocessing, of 23.65 $[ms]$, thus allowing to deploy it in real applications. The method was then tested and proved capable of compensating for undesired deformation of the carbon fiber ply both during the analysis of a step deformation response and in a manual guidance application.

Currently, the model is limited to movements in the three directions x-y-z, and it does not take into account rotations. Thus, we plan to acquire a dataset including also rotations. The proposed method uses, as input, depth images that are sensitive only to macroscopic deformations; thus, it is not particularly sensitive to traction forces that typically produce much lower intensity deformations. To reduce noise, depth images were averaged with the drawback of increased inaccuracy at higher robot tool speeds. To solve noisy inputs, the samples taken for each robot position during the dataset acquisition will increase, the camera noise during the data augmentation will

[1] Video describing the experiments available at https://zenodo.org/record/6379312

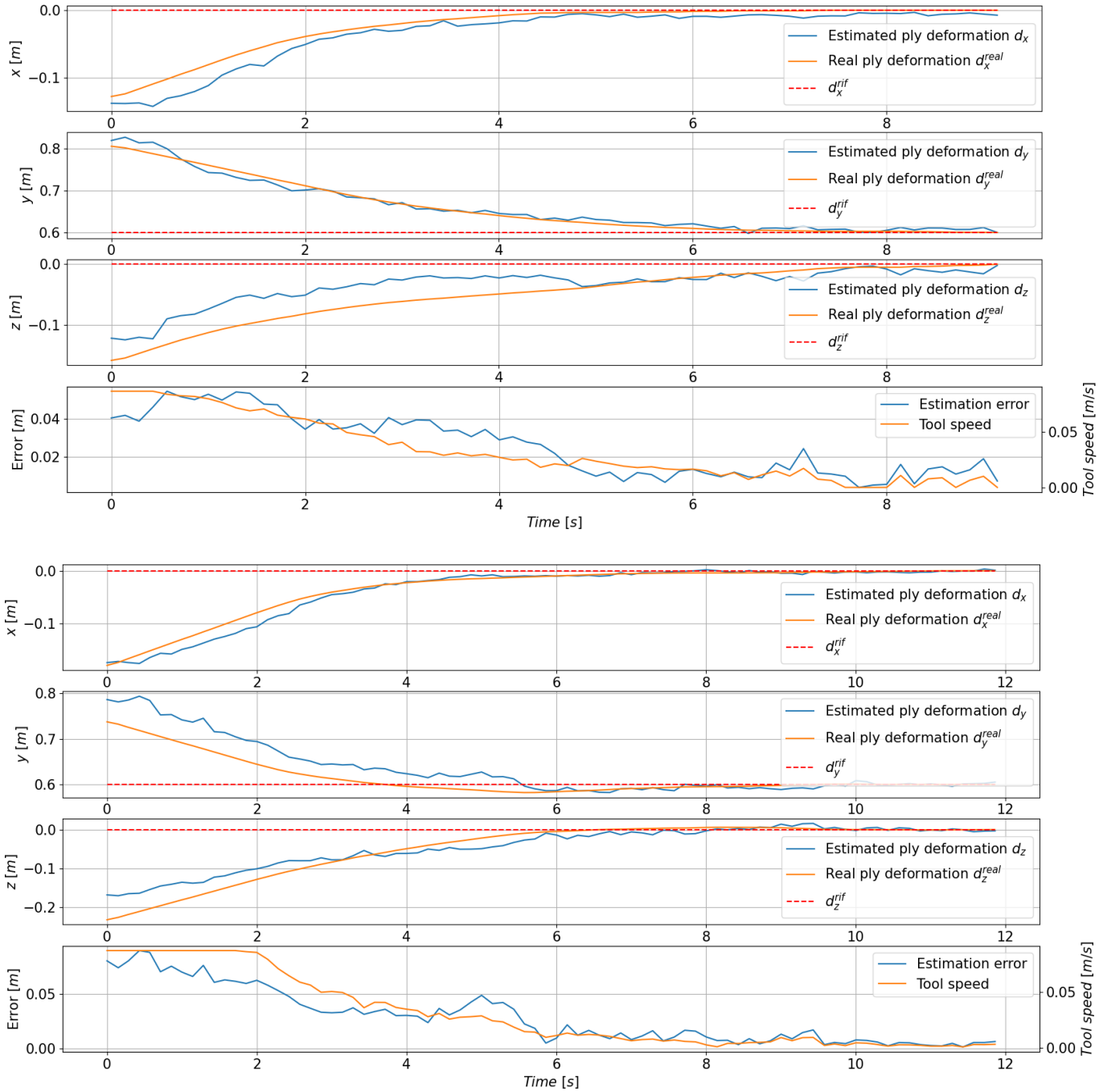

Fig. 3. Analysis of the step response and estimation error relation to tool speed. for each step response, (a) and (b), it shown the estimated and real ply deformation along the axis x-y-z and the total estimation error compared with the robot tool speed.

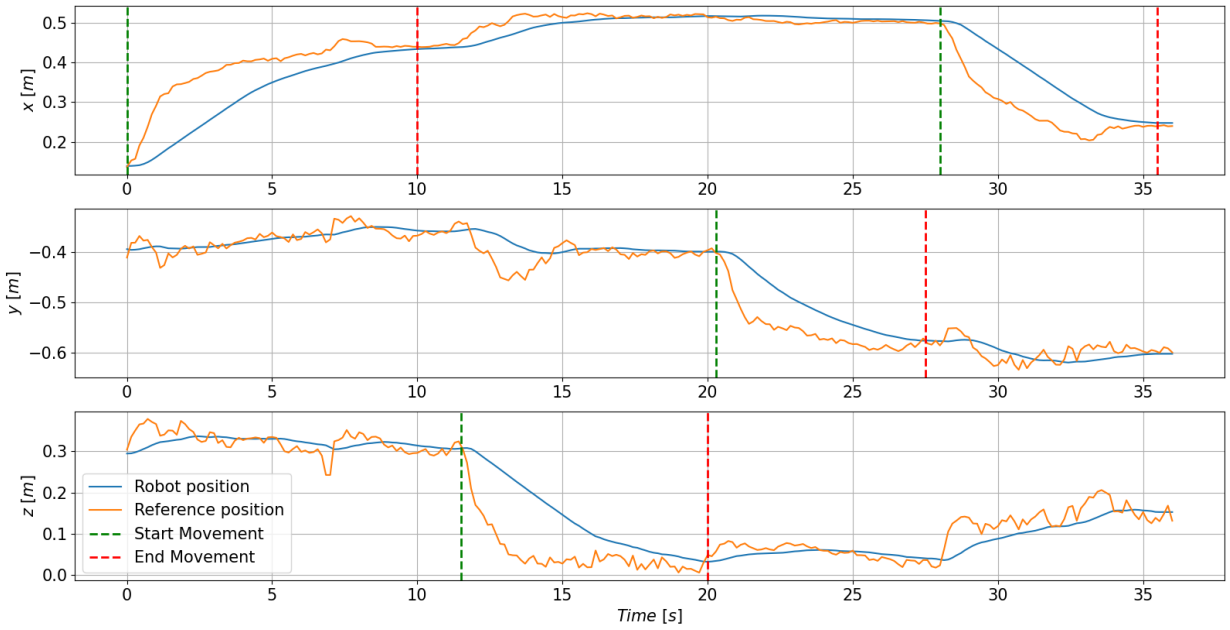

Fig. 4. Analysis of a manual guidance application. The human-robot reference displacement is converted into robot tool reference position (orange line). The robot position is the blue line. Green and red dashed lines highlight the start and end of successive human movement.

be simulated faithfully, and regularization during the training will be increased. Traction forces could be helpful to discriminate between movements that produce similar deformations like some translations and rotation. Therefore, we plan to introduce a sensor fusion between the forces and RGB-D images. Finally, the method was tested on a setup with a single industrial manipulator. However, introducing an IMM or a fleet of IMMs in a dynamic environment as a robotic partner would significantly increase the technological fallout of the work.

Acknowledgement

This paper has received funding from the EU's Horizon 2020 research and innovation program under grant agreement No 101006732, "DrapeBot – A European Project developing collaborative draping of carbon fiber parts".

References

[1] R. Herguedas, G. López-Nicolás, R. Aragüés, and C. Sagüés, "Survey on multi-robot manipulation of deformable objects," in *2019 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, 2019, pp. 977–984.

[2] D. De Schepper, B. Moyaers, G. Schouterden, K. Kellens, and E. Demeester, "Towards robust human-robot mobile co-manipulation for tasks involving the handling of non-rigid materials using sensor-fused force-torque, and skeleton tracking data," *Procedia CIRP*, vol. 97, pp. 325–330, 2021, 8th CIRP Conference of Assembly Technology and Systems.

[3] D. Kruse, R. J. Radke, and J. T. Wen, "Human-robot collaborative handling of highly deformable materials," in *2017 American Control Conference (ACC)*, 2017, pp. 1511–1516.

[4] D. Andronas, E. Kampourakis, K. Bakopoulou, C. Gkournelos, P. Angelakis, and S. Makris, "Model-based robot control for human-robot flexible material co-manipulation," in *2021 26th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, 2021, pp. 1–8.

[5] M. Aranda, J. Antonio Corrales Ramon, Y. Mezouar, A. Bartoli, and E. Özgür, "Monocular visual shape tracking and servoing for isometrically deforming objects," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 7542–7549.

[6] Y. Li, Y. Wang, Y. Yue, D. Xu, M. Case, S.-F. Chang, E. Grinspun, and P. K. Allen, "Model-driven feedforward prediction for manipulation of deformable objects," *IEEE Transactions on Automation Science and Engineering*, vol. 15, no. 4, pp. 1621–1638, 2018.

[7] D. Kruse, R. J. Radke, and J. T. Wen, "Collaborative human-robot manipulation of highly deformable materials," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*, 2015, pp. 3782–3787.

[8] B. Jia, Z. Hu, J. Pan, and D. Manocha, "Manipulating highly deformable materials using a visual feedback dictionary," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, 2018, pp. 239–246.

[9] B. Jia, Z. Pan, Z. Hu, J. Pan, and D. Manocha, "Cloth manipulation using random-forest-based imitation learning," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 2086–2093, 2019.

[10] Z. Hu, P. Sun, and J. Pan, "Three-dimensional deformable object manipulation using fast online gaussian process regression," *IEEE Robotics and Automation Letters*, vol. 3, no. 2, pp. 979–986, 2018.

[11] V. V. Vasiliev and E. V. Morozov, "Chapter 8 - equations of the applied theory of thin-walled composite structures," pp. 575–590, 2018.

[12] J. Wang and E. Olson, "AprilTag 2: Efficient and robust fiducial detection," in *Proceedings of the IEEE/RSJ International Conference on Intelligent, Robots and Systems (IROS)*, October 2016.

[13] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2015.

[14] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," 2019.