Tip Position Estimation of a Soft Inflatable Bellow Actuator: A Vision-Based Proprioceptive Sensing Approach

Matthias Hofer, Peter Werner, Carmelo Sferrazza and Raffaello D'Andrea

***Abstract*— We demonstrate a vision-based proprioceptive sensing approach for a soft inflatable actuator. The 3D tip position of a bellow actuator is predicted from the images taken by an internal camera. For this purpose, a supervised learning approach is employed, relying on ground truth data from a motion capture system. The elongation of the actuator is controlled by adjusting the internal pressure and the tip can be manually moved in an arbitrary lateral direction. The real-time prediction of the tip position has an accuracy of a few millimeters and is visualized on a laptop computer.**

I. Vision-based Sensing

Vision-based sensing that relies on an internal camera to provide soft robots with sensory feedback is a promising approach for two main reasons. Firstly, the sensor (i.e. the camera) is not required to match the compliance of the soft material employed and secondly, vision-based sensing approaches provide high spatial resolution (e.g. [1], [2]).

We demonstrate a vision-based sensing approach for a soft bellow-type actuator featuring an integrated camera. A distinctive white pattern is applied to the interior surface of the fabric, providing rich, visual information (see Fig. 1). A number of computationally cheap features are computed from the internal camera image and used in a supervised learning pipeline (based on support vector regression) to predict the 3D tip position of the actuator. A motion capture system provides ground truth data during the training.

The proposed approach does not suffer from pattern occlusion, which is inherent to the inflatable bellow actuator and hinders straightforward implementation of classical computer vision approaches. Moreover, the lightweight nature of the features employed facilitates real-time implementation and the support vector regression exhibits lower training complexity compared to end-to-end deep learning approaches.

The elongation of the actuator is adjusted by changing the internal pressure and the tip can be manually moved in an arbitrary lateral direction. The tip position is predicted in real time (40 Hz) with an accuracy of a few millimeters and visualized on a laptop computer connected to the test setup (see Fig. 2). Thereby, no ground truth is required for demonstration purposes. A detailed explanation of the approach can be found in [3] and a video is available under `https://youtu.be/ylBDW8_fC-A`.

II. Beyond Soft Robotics Vision

Vision-based sensing is a promising approach to provide soft robots with sensory information. The approach

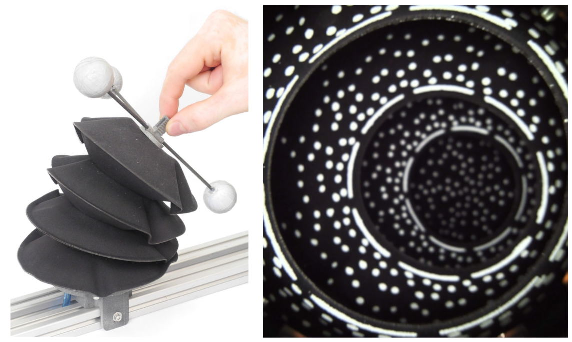

Fig. 1. Left: A linear soft inflatable bellow actuator with an integrated camera is manually deflected. Markers for an external motion capture system are mounted on top and are used as ground truth during training. Right: Image recorded by the internal camera showing the pattern employed in white. The proposed sensing approach predicts the tip position of the actuator (located on the grip) using only images from the internal camera.

demonstrated lies at the intersection of computer vision, machine learning and soft robotics. The rich visual information from soft materials undergoing deformation is captured with a camera providing high spatial resolution. The feature engineering before learning improves data efficiency and therefore permits easy adaptation of the approach to different soft robotic systems.

Furthermore, the principle sensing approach has the potential to provide more than only positional sensory feedback. Leveraging ideas from tactile sensing, the soft material forming the robot can be used in combination with vision sensors to provide shape information and interaction forces provided that ground truth data are available for training.

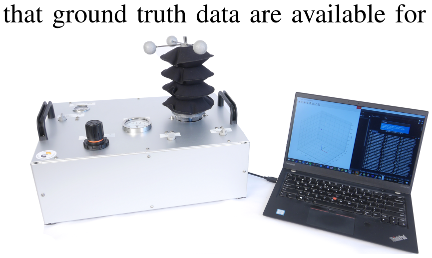

Fig. 2. The mobile demonstrator allows for adjustment of the pressure in the actuator and thereby changing its elongation. The camera-based prediction is executed on a laptop computer and visualized in real-time. The demonstrator is entirely self-sufficient and does not require any external peripherals such as electric power or pressurized air.

The authors are members of the Institute for Dynamic Systems and Control, ETH Zürich, Switzerland. Email correspondence to Matthias Hofer `hofermat@ethz.ch`.

References

[1] B. W. McInroe, C. L. Chen, K. Y. Goldberg, R. Bajcsy, and R. S. Fearing, "Towards a Soft Fingertip with Integrated Sensing and Actuation," in *IROS 2018*.

[2] R. Wang, S. Wang, S. Du, E. Xiao, W. Yuan, and C. Feng, "Real-time soft body 3d proprioception via deep vision-based sensing," *RAL 2020*.

[3] P. Werner, M. Hofer, C. Sferrazza, and R. D'Andrea, "Vision-based proprioceptive sensing for soft inflatable actuators," *arXiv preprint arXiv:1909.09096*, 2019.