2024 IEEE International Conference on Robotics and Automation (ICRA 2024)
May 13-17, 2024. Yokohama, Japan

Learning Motion Reconstruction from Demonstration via Multi-Modal Soft Tactile Sensing

Cheng Pan[1], Kieran Gilday[1], Emily Sologuren[2], Kai Junge[1] and Josie Hughes[1]

***Abstract*— Learning manipulation from demonstration is a key way for humans to teach complex tasks. However, this domain mainly focuses on kinetic teaching, and does not consider imitation of interaction forces which is essential for more contact rich tasks. We propose a framework that enables robotic imitation of contact from human demonstration using a wearable finger-tip sensor. By developing a multi-modal sensor (providing both force and contact location) and robotic collection of simple training data of different motion primitives (tapping, rotation and translation), an LSTM-based model can be used to replicate motion from tactile demonstration only. To evaluate this approach, we explore the performance on increasingly complex testing data generated by a robot, and also demonstrate the full pipeline from human demonstration via the sensor used as a wearable device. This approach of using tactile sensing as a means of inferring the required robot motion paves the way for imitation of more contact-rich tasks, and enables imitation of tasks where the demonstration and imitation is performed with different body-schema.**

I. Introduction

Learning from demonstration and imitation learning has been highly effective for advancing robotic manipulation capabilities, providing impressive task performance [1]. These approaches often rely on kinesthetic teaching, teleoperation or visual information [2], [3], utilising kinematic data from expert demonstration to learn skills rapidly [4]. However, for many tasks which are contact rich and leverage environmental interaction, it is important to imitate the contact and the physical interaction with the environment, opposed to simply the kinematics [1]. Kinesthetic teaching and teleoperation approaches often limit the expert demonstrators ability to teach the desired forces [5], and approaches that purely rely on visual information cannot guide force sensitive interactions [6]. To move towards teaching more adaptive and robust manipulation [7], [8], [9], we must consider how to capture and imitate contact from a human demonstrator.

Gaining and exploiting tactile information is hindered by limitations of existing sensor technologies and our limited understanding of how humans process tactile information for manipulation [10]. Soft sensors have proven to be essential for capturing environmental contact [10], [11], and typically measure mechanical deformation of a soft surface [12]. Sensing technologies include stretchable piezoresistive sensors [13], soft barometric sensors [14], acoustic sensing of body deformation [15] and more [16]. For imitation from human demonstration, we require a wearable system which can also be deployed on a robot. Although the development of soft sensors is broad, many lack this versatility [17]. A second challenge is that tactile information is sparse (spatially and temporally) meaning that accurately representing or reconstructing tactile responses during manipulation is challenging, and most often requires a data-driven approach [2], [18], [19]. For soft and compliant robots and sensors, this problem is even more apparent [13]. In addition to this challenge, when learning from indirect information, such as tactile data, we introduce an extra translation step when compared to learning purely from kinesthetic data [20], [21], [22]. Before we can apply knowledge from expert demonstrations, the robot must be able to accurately reconstruct motions and interactions given demonstration of sensor data [18], [21].

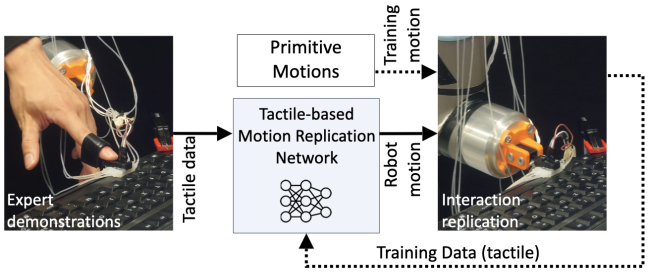


Fig. 1: Tactile-based motion replication is trained to generate robot motions from tactile information, enabling generation of open-loop trajectories without the need for kinesthetic teaching.

We hypothesise that a purely tactile-based motion replication framework can accurately reconstruct robot motions to generate the same tactile (time-sequential) trajectories as a human demonstrator. By developing a multi-axis, soft, force sensing array that can be worn by both human demonstrators and the robot reconstructor, we can capture and compare contact generated from human and robot actions. Secondly, we propose that by utilising only primitive motions to generate training data, and learning the relationship between sensor and robot motion, we can avoid the expensive step of directly learning replication of the tactile response. This approach has the potential to generalize across different task-environments by learning primitive interactions. These primitives can then be integrated together for more complex motions in downstream imitation learning tasks.

Our approach of reconstructing motion from contact is shown in Fig. 1. We have developed a soft sensor formed from an array of barometric pressure receptors and a 3-axis soft magnetic sensor. A long short-term memory (LSTM) network is trained on tactile data from primitive motions such as tapping, rotating or translating along a surface. This learnt

[1]CREATE Lab, Swiss Federal Institute of Technology Lausanne (EPFL), 1015 Lausanne, Switzerland. cheng.pan@epfl.ch
[2]Computer Science and Artificial Intelligence Lab, Massachusetts Institute of Technology (MIT), Cambridge, MA, USA

network can be used to generate more complex motions, for example: different combinations and frequencies of tapping, rotations in multiple axes simultaneously, and motions from human demonstrations. The framework is tested with different data modalities and across different environments. We see the barometric sensing array provides more useful information overall, but the additional shear sensitive modality of the magnetic sensor improves performance in sliding motions. During experiments, the robot is able to reproduce combinations of primitive motions and, despite errors, is able to reproduce tasks such as key presses in an environment unseen in the training data.

The results demonstrate that the robot successfully imitates the tactile response, hence interactions, from a range of human demonstrations with an average error of 9.8%. Therefore, we can rapidly and accurately imitate contact, regardless of the body schema of the demonstrator or robot within a controlled environment. To extend the generalizability, both to general motions and to different environments, further work is required to explore a wider set of training motion primitives, different modalities of sensors and integration with kinematic data.

In the remainder of this paper we first introduce the Methods and Materials including the sensor design and LSTM training methods. The Experimental Setup and Results following this before Section V presents the Discussions and Conclusions.

II. Methods & Materials

A. Sensor Design & Characterization

Inspired by the different receptors present in human skin and the requirement to capture force and spatial information, we have designed a multi-modal soft sensor. Our sensor consists of an array of barometric pressure sensors and a soft, magnetic field effect sensor (Fig. 2). It is constructed from a 3D printed base with an embedded hall effect sensor and a soft silicone layer cast from Ecoflex 00-20 which has a permanent magnet and eight air chambers embedded. Each of the air chambers (Fig. 2c) are connected via pneumatic tubes to barometric sensors (MPXH6115AC6U)(Fig. 2b) which report the pressure change from chamber deformation. The distribution of the eight sensors enables localization of contact forces. Additionally, when the soft skin is deformed, the embedded permanent magnet moves and the corresponding change in magnetic field is captured by the 3-axis hall effect sensor (TMAG5273) (Fig. 2a). This is especially valuable for measuring shear forces. Similar to human tactile sensors, our sensor has redundancy between the two modalities.

Fig. 3a shows typical sensor outputs for the soft magnetic sensor under normal and shear force. These experiments were performed by cyclically applying normal or shear forces via a robot arm equipped with a load cell and measuring the sensor readout. Due to the geometry of the silicone layer, the magnetic sensor is five times more sensitive to shear force than normal force. Significant hysteresis is seen in the measured field strength, necessitating time dependent networks such as LSTMs for data analysis.

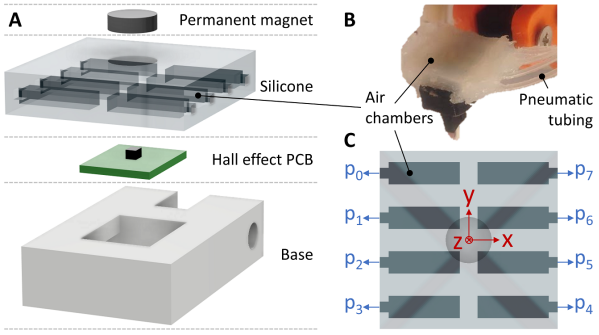


Fig. 2: Sensor design. (a) Exploded view. (b) Constructed sensor with pneumatic tubes connecting air chambers to pressure gauges. (c) Sensor schematic: the magnet can shift and rotate relative to the hall sensor in x, y and z; each air chamber connects to an independent pressure gauge.

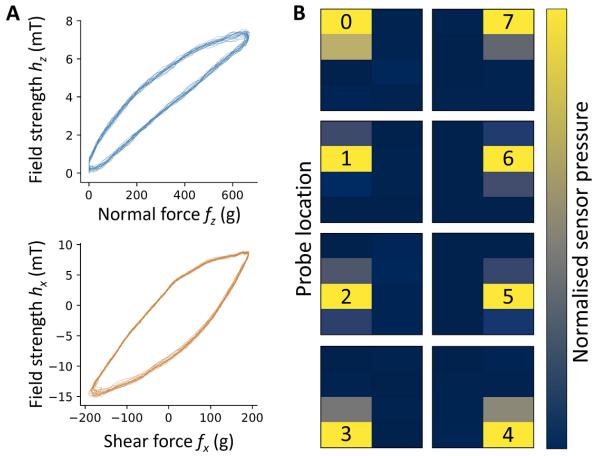


Fig. 3: Sensor characteristics. (a) hall effect sensor dominant axis signals under normal (z) and shear forces (x). (b) pressure sensor tactile maps under 650 g force in eight locations centered over each air chamber.

To show the localization capabilities of the pressure sensor array, the normalized pressure readings are shown when probed at the eight different sensor location with a fixed load (650 g) using a small indenter mounted on a robot arm (Fig. 3b). The characterization results show that the measured pressure can localize contact with high selectivity.

B. LSTM-based Trajectory Learning

The learning task of the network is to build the mapping between input tactile sensor signals x_t and motion trajectories y_t. The network for motion replication consists of one LSTM network and two fully connected layers, all hidden layers have 10 hidden nodes. The LSTM is trained on data collected by the robot where the robot motion is selected and the sensor readout is recorded.

Due to the robot's internal controller, the sensor signal x_t is actually triggered by motion y_{t-t_d}. As such, a typical LSTM network could only learn the mapping between x_t and y_{t-t_d} instead of desired x_t and target motion y_t. To address this issue, we concatenate the current sensor signal x_t with future information, three time steps ahead, i.e., $\{x_t, x_{t+t_a}, x_{t+2t_a}, x_{t+3t_a}\}$, where t_a controls the step length

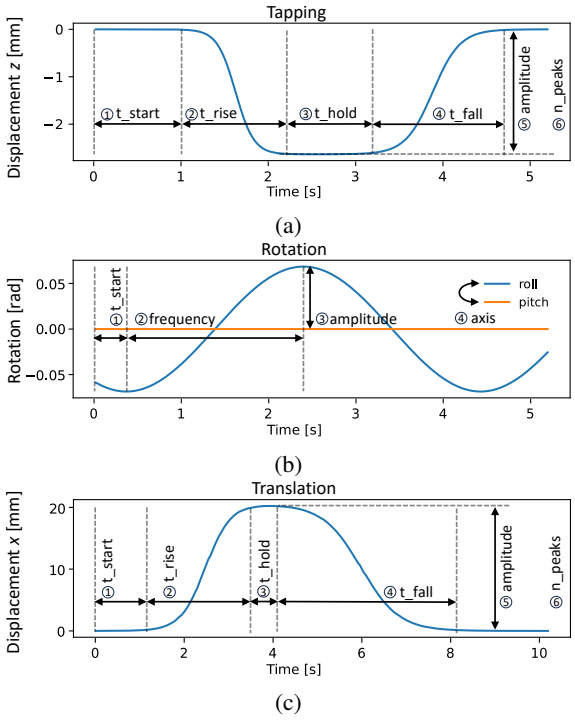


Fig. 4: Parameterised motion primitives. Tapping in the z direction (a) and translation in the x direction (c) both have 6 parameters, rotation in the roll (r_y) and pitch (r_x) axes (b) has 4 parameters.

ahead, and t is the current time for the sampled future information.

Due to the softness of the sensor, when undergoing a sliding motion, the physical interpretation (displacements) can be considered as an integration of the sensor readings over time. The hidden states of LSTM store past sensor information, providing similar information to integration of the sensor signals. However, to enhance the learning performance, we also provide the integrated tactile signals as an input to the network. As the LSTM lacks any differential information, differential signals have also been introduced as an input to the network for comparison purposes.

C. Training & Testing Data

Three different sets of training data are used, generated from different primitive robot motions as parameterised in Fig. 4. Tapping and translation trajectories (Fig. 4a, c) are identical, except for magnitudes and motion axes, and are parameterized by smoothed trapezoidal waves with controlled delays, rise times and amplitudes. Rotation trajectories (Fig. 4b) are generated by parametric sine waves in the r_x or r_y axes of the end-effector of the robot (Fig. 2c).

The training data for tapping and translation (one degree of freedom primitives) consists only of trajectories with a single peak, while the testing trajectories can have multiple peaks. For rotation (two degree of freedom), the training trajectories are given in only a single axis at a time — roll or pitch — whereas testing data can be in both axes simultaneously with higher frequencies.

For each form of motion primitive, we collect 100 trajectories as training data and 20 trajectories for validation

TABLE I: Range of amplitudes (relative to a fixed starting point a surface) of the different datasets (training and robot test) for the different motion primitives in Fig. 4.

	Train	Test (low)	Test (med)	Test (high)
Tapping [mm]	[2.2, 3.0]	[2.0, 2.4]	[2.4, 2.6]	[2.8, 3.2]
Rotation [rad]	[0.05, 0.09]	[0.04, 0.06]	[0.06, 0.08]	[0.08, 0.10]
Translation [mm]	[10, 26]	[6,14]	[14, 22]	[22, 30]

data. The amplitude of the motion primitives are selected randomly within the bounds shown in Table I. In addition to the training data, we capture robot generated testing data, where the robot is used to collect sensor data for a given motion. This test data includes motion trajectories with variable amplitudes to reflect varying force application (low, medium and high amplitudes relative to the training data) and also variable frequencies or number of peaks (one, two or three peaks). 25 trajectories are generated for each category and their full description is given in Table I. The duration of all the tapping and rotation trajectories is 5 seconds, whilst for the translation they are 10 seconds. All training and testing data is collected from motions on a flat and solid surface. The robot finger always starts a fixed distance from the surface.

D. Error Metrics

To measure the performance of the reconstructed contact and motion we use various error metrics. The error metric for contact is the normalized sum of error between the demonstrated (x) and reconstructed sensor signals ($\hat{x}$) across all sensors (pressure sensor: $x_t^{p_i}$, hall sensor $x_t^{h_j}$), i.e.:

$$e_s^p = \frac{1}{T}\sum_{t=0}^{T-1}\left(\frac{1}{8}\sum_{i=0}^{7}\left|x_t^{p_i} - \hat{x}_t{}^{p_i}\right|\right) \tag{1}$$

$$e_s^h = \frac{1}{T}\sum_{t=0}^{T-1}\left(\frac{1}{3}\sum_{j=0}^{2}\left|x_t^{h_j} - \hat{x}_t{}^{h_j}\right|\right) \tag{2}$$

$$e_s^{norm} = \frac{1}{2}(e_s^{p,norm} + e_s^{h,norm}) \tag{3}$$

This error metric is valid when using either robot generated or human generated sensor demonstrations. When training data is generated via the robot arm, we also have access to the ground-truth motion y, thus we can also calculate an error metric in the motion space:

$$e_m = \frac{1}{T}\sum_{t=0}^{T-1}(y_t - \hat{y}_t) \tag{4}$$

For both metrics, the error is normalized based upon the maximum and minimum value of all sensor signals or motion trajectories in the training dataset to enable meaningful comparison. The normalization of the error of pressure and hall sensors is performed separately based on their respective maximum and minimum values.

III. Experimental Setup

The experimental setup in Fig. 5 was developed to collect the training and test data, and also for performing the

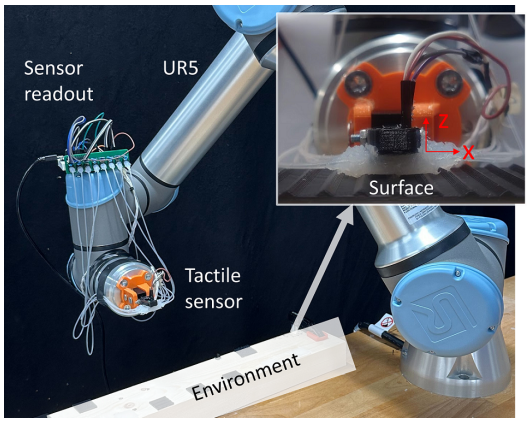


Fig. 5: Experimental setup. The tactile sensor can be driven to interact with the environment by a robot arm. The environment can be changed depending on the task.

reconstructed motions. The sensor is placed on a UR5 robot arm and gathers training data by probing the environment with the desired motion. A PC is used to simultaneously control the robot arm and read the sensor values from a microcontroller via serial communication. The hall-effect sensor data is obtained over I2C, and the eight pressure sensors are read via analog inputs on the 10-bit Arduino ADC. The sensors are sampled at 31.5 Hz.

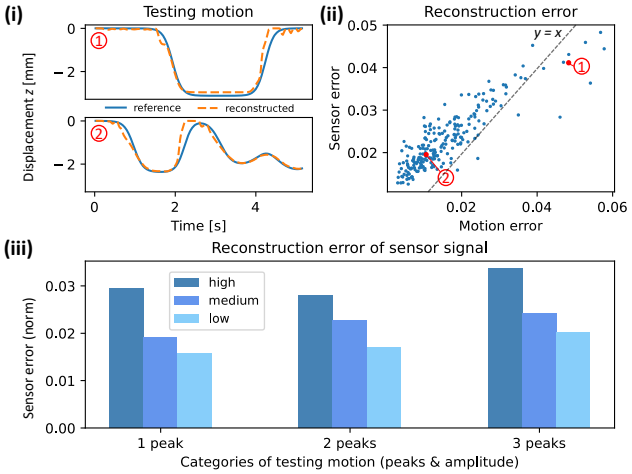
To first validate the pipeline, robot testing data was collected from motions performed on a flat wooden surface. The trained LSTM is then used to reconstruct the motion when provided with the sensor data collected by the robot. Following this, experiments are performed where we reconstruct motion directly from human demonstrations, where the sensor is used in its wearable format on the finger of the expert.

IV. Experimental Results

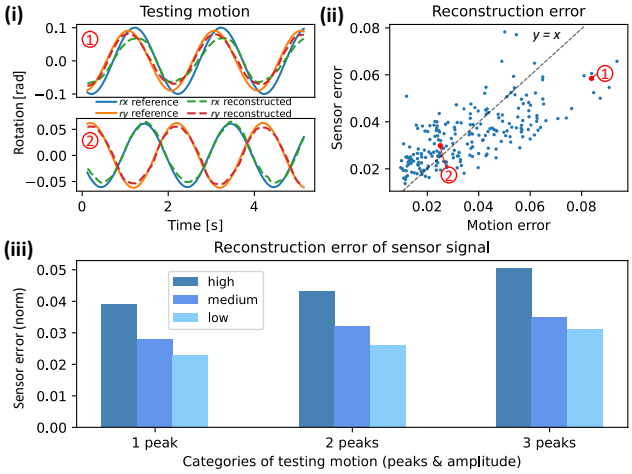
A. Motion Reconstruction

1) Tapping: Fig. 6a shows the reconstruction error for tapping. The relationship between motion error and sensor error has a linear and strongly correlated relationship. This suggests that minimizing the motion error also minimizes the error in the sensor space. Therefore, we can use error in motion error as a meaningful estimation of the sensor error. Examples of testing motion (Fig. 6ai) show that the reconstruction can generalize the single-peak training motion to more complex motions with variable amplitudes and multiple-peak motions. For different categories of testing motion with variable amplitudes and frequencies, we see that the largest error is always with higher amplitude motion and hence higher force contact. Although the error is similar across different number of peaks, it is highest for 3 peaks, which is most dissimilar to the training data, and where the sensor signal shows the greatest complexity.

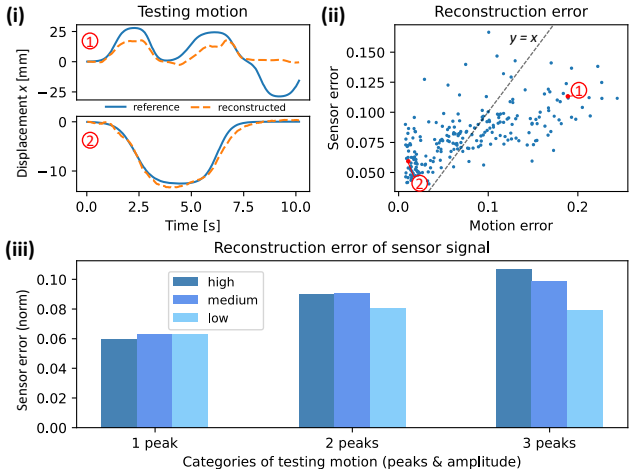
2) Rotation: Similar analysis is given for rotation motions in Fig. 6b. For rotational movements, the relationship between sensor error and motion shows that although the motion error can be high, the sensor error can be low. However, the motion error remains a good indicator of



(a)



(b)



(c)

Fig. 6: Motion reconstruction results for tapping (a), rotation (b) and translation (c). Example testing motions, in (i), constructed from motion primitives (Fig. 4. Sensor error vs. motion error of reconstructed motions, in (ii), we expect sensor error to increase with motion error. Average sensor error of reconstructed motions using different testing sets, in (iii). Testing sets categorised by motion amplitude (definition see Table I) and number of primitives they consist of (peaks).

TABLE II: Motion error (mean over time and 10 trials) using different input information, the range of motion is tapping: $[0,3]$ mm, rotation: $[-5.16,5.16]$ deg, translation: $[-25,25]$ mm

Motion	Standard	Diff.	Int.	Future	Future+diff.	Future+int.
Tapping [*mm*]	0.14	0.06	0.14	**0.05**	**0.05**	0.06
Rotation [*deg*]	0.57	**0.34**	0.52	0.37	0.43	0.36
Translation [*mm*]	5.60	4.38	4.01	4.05	3.86	**3.68**

sensor error in most cases. The example reconstructed testing motions show that the phase and temporal variation is well captured. The greatest error is seen reaching high amplitude peaks. This is similar to tapping, with increasing error at higher amplitudes and the number of peaks.

3) Translation: For translation motions (Fig. 6c) there is the weakest correlation between sensor error and motion error. The example trajectories show that the error increases towards the end of the trajectory. Additionally, while changes in direction are identified, the magnitude is not always well reconstructed. Unlike the other two motions, the error increases significantly for testing motions with more peaks, though the error has less of a dependency on motion amplitude. This could also reflect the physical interactions between the sensor and the environment. The readings of the pressure and hall sensors rely on the deformation of the silicone skin. This deformation remains after the sliding motion has finished and only disappears when the direction is reversed. Therefore, detection of the stop point relies on observing the change from kinetic to static friction. This can be challenging when only using tactile signals, and translation with multiple direction changes can cause larger error.

Overall, we see a mean sensor error across all categories as 2.3% for tapping, 3.4% for rotation and 8.1% for translation. This high accuracy suggests the trained networks generalize well to arbitrary combinations of primitive motions within the training environment.

B. Dependency on Future, Differential, and Integral Signal

Table II reports the reconstruction error in motion using different input-modes to the LSTM, i.e., with or without additional future, differential, or integral information (as described in Section II-B). In general, future information enhances the motion reconstruction performance in all motion primitives, as it can mitigate the time lag issue in training dataset. Without future information, differential information is also helpful. Additional integral information is especially important for motion reconstruction of translation.

1) Future information: When the future information is not included, the reconstructed motion in Fig. 7 (Standard) shows a time lag, leading to a higher reconstruction error (see Table II). In contrast, the time delay issue is mitigated when training with future information, which improves the learning performance for all motions when compared to the standard training inputs.

2) Differential information: Table II shows that with additional differential inputs, the reconstruction error can also be reduced and is comparable to that with future information. The differential information can reveal the changing trends in the sensor data, thereby facilitating faster response

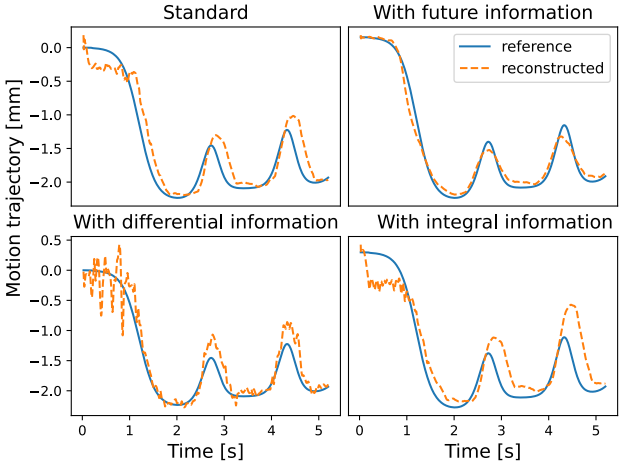


Fig. 7: Example reconstructed tapping motions with different input information.

in the motion reconstruction. This indicates that for some prediction tasks where future information is not available, using the differential of the input sensor signal can also help to mitigate the time lag problems.

3) Integral information: Incorporating integral information as an input to the LSTM does not significantly contribute to the accuracy of the reconstructed tapping and rotation motions. As the LSTM stores integral information in the hidden states, it already has information which is analogous to signal integration. However, for the translation motion, incorporating integration information is advantageous. This can be explained by the design of the task. For translation, the target trajectories is displacement in the x-axis, which can be regarded as the integration of the extent of deformation over time. Hence, the output motion has a strong correlation to the integral of the sensor data. Therefore, additional integral information enhances the learning performance.

4) Combination: The combination of future information with differential or integral information further reduces the reconstruction errors for translation motions, as the relationship between displacement in x-axis and sensor readings is more complex than for tapping and rotation.

Overall, future information without differential or integral information performs well across all motions and thus remaining experiments use this type of data input.

C. Sensor Dependency

To explore the contributions of the different sensor modalities to the motion reconstruction, the LSTM was trained using different combinations of sensor data: pressure sensor only, hall effect only, and both sensors. For the tapping tasks, the motion reconstruction error was similar for all sensor combinations, with pressure only showing a slight advantage (Fig. 8). For rotation tasks, the combined pressure and hall effect sensor data performed best. However, the hall effect only sensor data had significantly worse performance (almost 5 times the error). This indicates that spatial information (provided by the pressure sensors) is beneficial for reconstruction of rotation motions. For the translation, where

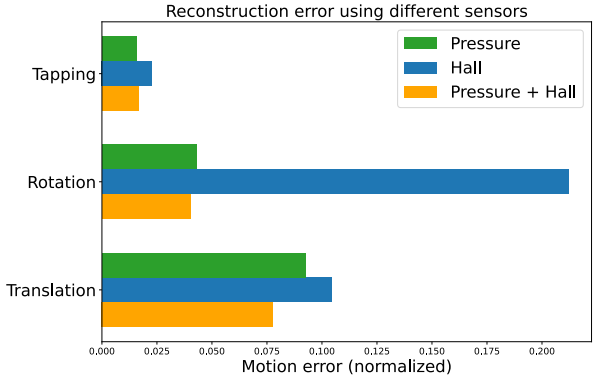


Fig. 8: Motion error when using different sensor type combinations for each motion type.

normal and shear forces are applied, the lowest error is also achieved with a combination of pressure and hall effect. These results demonstrate that multi-modality in sensing can be beneficial even when there is some redundancy in information between the modalities.

D. Human Demonstrations

To demonstrate the full pipeline (Fig. 1) of motion reconstruction from human demonstrations, we test the tapping, rotation, and translation motion separately using the wearable tactile sensor on the flat surface. Table III shows the sensor error for the first imitation tasks across 10 examples per each motion type. Compared to the results using the testing dataset, where the motion is executed by robot arm, the overall mean error increases from 4.6% to 9.8%. This is likely caused by the presence of small, high frequency vibrations in human motions. Despite the significant increase, the reconstructed motion retains sufficient accuracy to complete many manipulation tasks [2], [23].

We also consider a demonstration 'task' which combines different motions including tapping and sliding the demonstrator's fingertip over a keyboard. Fig. 9 shows images of the demonstrated manipulation task which includes two taps on one key, translation to another key and three taps afterwards. We first split the tactile signal for different neural networks to generate corresponding motions. The reconstructed movement of robot arm shows similarity to the human motion, which is demonstrated through successful key presses that match the humans. The error in sensor signal varies over time and the highest amplitude error appears during the tapping motion. The high error value occurs when the key is moving and traveling into the keyboard. When the tapping motion is executed, the key does not provide the same force feedback as in training, when a solid surface was used. The keyboard environment is more complex and dynamic than that used for training, and has different surface properties. Despite this the LSTM trained on a flat surface provides the ability to replicate motion and trigger key presses.

V. Discussion & Conclusion

In this paper, we present a framework for reconstruction of complex motions via human demonstration using a wearable multi-modality sensorised finger tip. The reconstruction leverages a trained LSTM which utilised robot gathered training data on primitive motion strategies. We demonstrate the ability to generalize and reconstruct more complex motions (e.g. motions described by combinations of primitives and noisy human demonstrations) from simple motions (e.g. trapezoids and sinusoids). The contributions of the multi-modality sensor are shown, with the combination of two sensors offering the best training reconstruction over the three different motion primitives explored.

Our tactile-based motion reconstruction approach utilizes sensing data as a means of grounding human and robot action, enabling imitation of a motion from demonstration of contact. This approach is particularly useful for interactions which are contact rich and also where the body schema between the demonstrator and replicator are different. Further work is required to move towards reconstruction of more generalized motions. Secondly, the current method only enables generation of open-loop trajectories which cannot accurately reconstruct interactions on dynamic surfaces without prior information about the surface. Augmenting this method with other information, e.g. visual or kinematic, or as part of a feedback controller could enable generalization in more varied environments.

TABLE III: Sensor error from trajectories reconstructed from human demos with the wearable fingertip (mean from 10 demonstrations and reconstructions).

	Pressure [%]	Hall [%]	Mean [%]
Tapping	6.07	8.64	7.36
Rotation	6.78	10.01	8.40
Translation	14.15	13.33	13.74

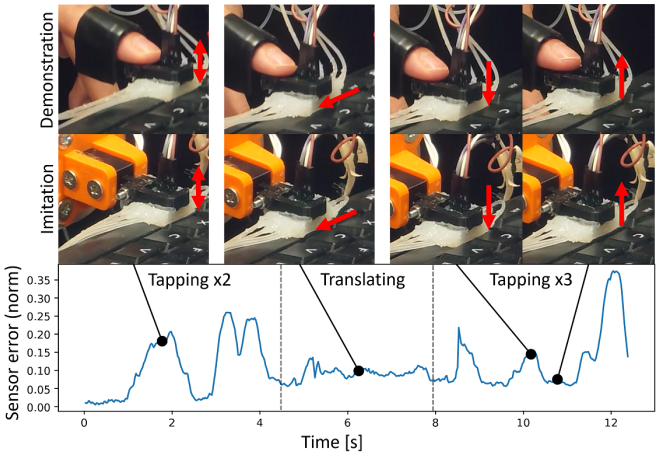


Fig. 9: Imitation learning pipeline example. Mean sensor error from the reconstruction of a tapping-translation-tapping motion on a keyboard. Significant errors are seen when the environment is able to deform under applied force.

Acknowledgements

This project was supported by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 945363.

REFERENCES

[1] B. Fang, S. Jia, D. Guo, M. Xu, S. Wen, and F. Sun, "Survey of imitation learning for robotic manipulation," *International Journal of Intelligent Robotics and Applications*, vol. 3, pp. 362–369, 2019.

[2] J. Hua, L. Zeng, G. Li, and Z. Ju, "Learning for a robot: Deep reinforcement learning, imitation learning, transfer learning," *Sensors*, vol. 21, no. 4, p. 1278, 2021.

[3] R. Memmesheimer, I. Kramer, V. Seib, and D. Paulus, "Simitate: A hybrid imitation learning benchmark," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 5243–5249.

[4] Y. Duan, M. Andrychowicz, B. Stadie, O. Jonathan Ho, J. Schneider, I. Sutskever, P. Abbeel, and W. Zaremba, "One-shot imitation learning," *Advances in neural information processing systems*, vol. 30, 2017.

[5] H. Kim, Y. Ohmura, A. Nagakubo, and Y. Kuniyoshi, "Training robots without robots: deep imitation learning for master-to-robot policy transfer," *IEEE Robotics and Automation Letters*, vol. 8, no. 5, pp. 2906–2913, 2023.

[6] J. J. Steil, F. Röthling, R. Haschke, and H. Ritter, "Situated robot learning for multi-modal instruction and imitation of grasping," *Robotics and autonomous systems*, vol. 47, no. 2-3, pp. 129–141, 2004.

[7] K. Li, D. Chappell, and N. Rojas, "Immersive demonstrations are the key to imitation learning," *arXiv preprint arXiv:2301.09157*, 2023.

[8] S. Tian, F. Ebert, D. Jayaraman, M. Mudigonda, C. Finn, R. Calandra, and S. Levine, "Manipulation by feel: Touch-based control with deep predictive models," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 818–824.

[9] Z. Kappassov, J.-A. Corrales, and V. Perdereau, "Touch driven controller and tactile features for physical interactions," *Robotics and Autonomous Systems*, vol. 123, p. 103332, 2020.

[10] B. Shih, D. Shah, J. Li, T. G. Thuruthel, Y.-L. Park, F. Iida, Z. Bao, R. Kramer-Bottiglio, and M. T. Tolley, "Electronic skins and machine learning for intelligent soft robots," *Science Robotics*, vol. 5, no. 41, p. eaaz9239, 2020.

[11] K. Gilday, T. George-Thuruthel, and F. Iida, "Predictive learning of error recovery with a sensorized passivity-based soft anthropomorphic hand," *Advanced Intelligent Systems*, p. 2200390, 2023.

[12] H. Wang, M. Totaro, and L. Beccai, "Toward perceptive soft robots: Progress and challenges," *Advanced Science*, vol. 5, no. 9, p. 1800541, 2018.

[13] T. G. Thuruthel, K. Gilday, and F. Iida, "Drift-free latent space representation for soft strain sensors," in *2020 3rd IEEE International Conference on Soft Robotics (RoboSoft)*. IEEE, 2020, pp. 138–143.

[14] K. Gilday, L. Relandeau, and F. Iida, "Design and characterisation of a soft barometric sensing skin for robotic manipulation," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 10 021–10 028.

[15] V. Wall, G. Zöller, and O. Brock, "Passive and active acoustic sensing for soft pneumatic actuators," *The International Journal of Robotics Research*, vol. 42, no. 3, pp. 108–122, 2023.

[16] Y. Luo, X. Xiao, J. Chen, Q. Li, and H. Fu, "Machine-learning-assisted recognition on bioinspired soft sensor arrays," *ACS nano*, vol. 16, no. 4, pp. 6734–6743, 2022.

[17] H. Kristanto, P. Sathe, A. Schmitz, T. P. Tomo, S. Somlor, and S. Sugano, "A wearable three-axis tactile sensor for human fingertips," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 4313–4320, 2018.

[18] D. Xu, S. Nair, Y. Zhu, J. Gao, A. Garg, L. Fei-Fei, and S. Savarese, "Neural task programming: Learning to generalize across hierarchical tasks," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 3795–3802.

[19] J. Hughes, A. Spielberg, M. Chounlakone, G. Chang, W. Matusik, and D. Rus, "A simple, inexpensive, wearable glove with hybrid resistive-pressure sensors for computational sensing, proprioception, and task identification," *Advanced Intelligent Systems*, vol. 2, no. 6, p. 2000002, 2020.

[20] P. Kormushev, S. Calinon, and D. G. Caldwell, "Imitation learning of positional and force skills demonstrated via kinesthetic teaching and haptic input," *Advanced Robotics*, vol. 25, no. 5, pp. 581–603, 2011.

[21] H. Liu, C. Zhang, Y. Zhu, C. Jiang, and S.-C. Zhu, "Mirroring without overimitation: Learning functionally equivalent manipulation actions," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, 2019, pp. 8025–8033.

[22] M. Edmonds, F. Gao, X. Xie, H. Liu, S. Qi, Y. Zhu, B. Rothrock, and S.-C. Zhu, "Feeling the force: Integrating force and pose for fluent discovery through imitation learning to open medicine bottles," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017, pp. 3530–3537.

[23] S. Scherzinger, A. Roennau, and R. Dillmann, "Contact skill imitation learning for robot-independent assembly programming," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 4309–4316.