AllSight: Advancing Optical Tactile Sensing and Sim-to-Real Learning for Dexterous Robotic Manipulation

Osher Azulay and Avishai Sintov

***Abstract*— In this work, we introduce *AllSight*, a novel optical tactile sensor with a round 3D structure designed for robotic in-hand manipulation. We highlight its primarily 3D-printed construction which offers affordability, modularity, durability and a human thumb-sized design with a large contact surface. With the design, the ability to learn and estimate a full contact state, i.e., contact position, forces and torsion, is shown through experimental benchmarks. Next, to tackle the reality gap in simulators for high-resolution tactile sensors like AllSight, we propose *SightGAN*, a bi-directional Generative Adversarial Network that refines the sim-to-real transition, particularly for 3D round sensors, and facilitates the training of zero-shot models for newly fabricated sensors. This approach not only advances the potential for dexterous robotic manipulation but also showcases the significant improvement in creating realistic synthetic images.**

I. Introduction

Tactile sensing plays a pivotal role in human perception, making it a significant area of exploration in robotics research [1]–[3]. It is essential for providing robots with the ability to interact with the environment through precise and dexterous actions. While various compact sensors with flat contact surfaces have been developed [4]–[6], these often encounter challenges in complex manipulation tasks due to their surface geometry. Consequently, sensors with spatially varied surface geometries were introduced [7], [8]. Yet, these sensors struggle to deliver comprehensive and reliable contact information. Limitations include incomplete contact state data [9], restrictions in load sensing [1], or the need for simple and low-cost manufacturing processes [10].

In this work, we introduce a framework for optical-based tactile sensors which includes hardware, data-based modeling and realistic simulator. First, we propose *AllSight*, a compact and cost-effective solution for multi-fingered robotic hands engaged in in-hand manipulation tasks. Its 3D contact surface is shaped as a cylinder with a hemispherical end as illustrated in Figure 1.A. Moreover, pre-trained on simulation and real contact datasets, AllSight is capable of providing precise full contact state information including position, normal and tangential forces, and torsion on newly fabricated sensors.

Furthermore, the advancement in tactile sensor technology has led to the need for complex data representations, pushing for large-scale datasets to develop accurate models [11]. Addressing the demand for extensive data, simulations for optical-based tactile sensors have been designed to quickly generate vast tactile image datasets [12], [13]. Yet, bridging the gap between simulated and actual sensor data poses challenges due to discrepancies in tactile images [14], [15]. To tackle the sim-to-real challenge we propose the *SightGAN* model which augments CycleGAN [16] with specific consistency losses. These losses aim at reducing background differences and improving contact position accuracy between simulated and real tactile images. Its bidirectional capability ensures effective knowledge transfer across simulated and real settings, optimizing model training despite variations in sensor illumination and design.

O. Azulay and A. Sintov are with the School of Mechanical Engineering, Tel-Aviv University, Israel. E-mail: osherazulay@mail.tau.ac.il, sintov1@tauex.tau.ac.il

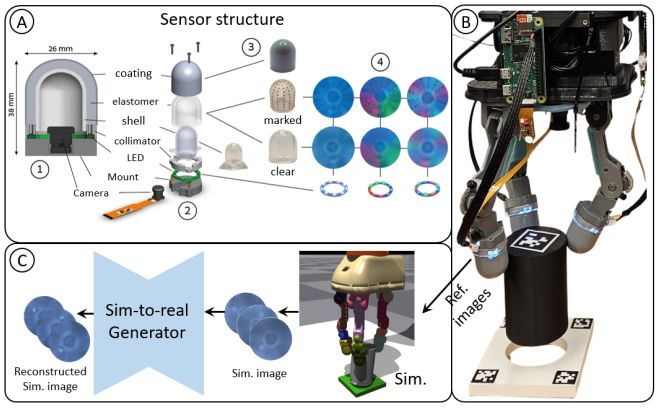

Fig. 1: Illustration of AllSight. (A) Sensor structure, (B) Three AllSight sensors on the fingers of an adaptive hand. (C) sim2real generator from the trained SightGAN model is used to map simulated tactile images to real-like images of a 3D round tactile sensor

II. Method

A. Design

The AllSight sensor is an optical-based tactile sensor designed for comprehensive 360° contact sensing without any blind spots. It features a cylindrical tube with a hemispherical end, as illustrated in Figure 1.A. The tube consists of a three-layered structure. At its core is a rigid, transparent shell fabricated through SLA 3D printing. The shell is enveloped by a layer of transparent elastomer and finally coated with reflective silicone paint. A camera positioned inside the tube captures the deformation of the elastomer layer when it comes into contact with an object. Illumination inside the shell is provided by LEDs arranged on an annular PCB, ensuring clear and detailed image capture. The LED system can support various lighting settings.

B. Data collection

The contact state of AllSight includes the contact position, force and torsion. Data is collected by sampling real and simulated labeled image datasets. In the collection of real images, each image is labeled using a robotic arm equipped

with a Force/Torque (F/T) sensor and an indenter, allowing for precise control over contact location and pressure. The dataset pairs each image $\mathbf{I}_i$ with a state measurement, detailing position calculated via the arm's forward kinematics and the contact load. In a simulated dataset, we employ TACTO [12], a physics-engine simulator tailored for optical-based tactile sensors. This process involves calibrating a virtual AllSight sensor with reference images from actual sensors to enhance sim-to-real transfer accuracy. Gaussian noise and various illumination settings were added to augment the simulated images. Each sampled simulated image is labeled with the contact position.

C. Contact estimation model

A state estimation model infers the contact state from a tactile image. The model is a modified ResNet-18 model, removing the top layer and incorporating two fully-connected layers with ReLU activations. At each iteration, both reference $\mathbf{I}_{ref}$ and contact $\mathbf{I}_i$ images are down-sampled to resolution 224×224 and stacked along the channel. Furthermore, we also consider difference images in the dataset such that an image used for training is $\hat{\mathbf{I}}_i = \mathbf{I}_i - \mathbf{I}_{ref}$ in order to make the model agnostic to the background and focuses only on the color gradients that occur around the deformations. Utilizing simulated data simplifies initial training, requiring less real data for model refinement. Hence, the contact encoder, pre-trained on simulated dataset, is fine-tuned with real-world data to enhance accuracy.

D. Sim-to-real with SightGAN

SightGAN extends the CycleGAN [16] framework to address the challenges in tactile image translation between simulated and real domains. It incorporates specialized auxiliary losses to enhance the translation fidelity, particularly focusing on tactile perception nuances. The architecture employs the standard CycleGAN cycle consistency loss $\mathcal{L}_{\text{cycle}}$ for preserving image integrity through domain translations. To specifically support the tactile image domain, SightGAN introduces auxiliary losses designed to ensure accurate contact localization and maintain structural integrity in the tactile images in both spatial and image domains. For the spatial domain, we define the Spatial Contact Consistency loss following the same structure of the consistency loss in [17] where we replace the perception function with

$$\mathcal{L}_{sp}(\mathbf{I}_i, \mathbf{I}_j)) = \|f_\theta(\mathbf{I}_i) - f_\theta(\mathbf{I}_j)\|^2 \quad (1)$$

where f_θ is the contact position estimation function. To further augment the accuracy of contact localization in domain transfer and enhance structural fidelity, we introduce a loss related to the contact region. For each image $\mathbf{I}$, binary image $\mathbf{B}$ is defined where a mask is placed on the contact region of the image. The contact loss between an image and its transfer is, therefore, defined by

$$\mathcal{L}_{\text{m}}(\mathbf{I}, \mathbf{B}, H) = \|\mathbf{I} * \mathbf{B} - H(\mathbf{I}) * \mathbf{B}\|_1 \quad (2)$$

where H is the generator of CycleGAN. The overall loss function for SightGAN integrates these components to optimize the tactile image translation process, enabling more accurate and reliable sim-to-real and real-to-sim translations.

TABLE I: Estimation accuracy of contact positions

	Origin of training data for f_θ	Position RMSE (mm)
Direct	Data from 6 train sensors	2.16
	6 sensors from simulation	7.48
Gen.	CycleGAN	13.30
	SightGAN	3.49

III. Experimental Results & Discussion

We first evaluate the precision of state estimation using collected data. First, the model was trained with 20,000 train images from a real sensor with white illumination featuring a single spherical indenter of 3 mm radius. Evaluated over 1,282 test images, the yielded mean position, force and torsion are 0.79±0.27 mm, 0.9±0.41 N and 0.002±0.001 Nm, respectively.

Furthermore, we analyse the position estimation with SigntGAN. Table I summarizes the Root-Mean-Square-Errors (RMSE) for position estimation with model f_θ while trained with different origins of training data. The results include accuracy when training directly with data from six real sensors. Next, model f_θ is trained with data generated in the simulation without any GAN and while using the reference images of the six training sensors. Using only simulated data provides poor accuracy showing that the simulation, even with real reference images, is far from representing reality. Then, f_θ with data generated by the CycleGAN alone without additional losses is evaluated. The error with only CycleGAN is the highest due to its inability to focus and reconstruct the contact. Adding the two losses of SightGAN significantly reduces the error. Next, we evaluate the accuracy which the sim-to-real of SightGAN provides to images generated from simulation. Figure 2 shows the error of position estimation over the test data of the two new sensors with regards to the number of new samples used to fine-tune the model. With no additional data, i.e., zero-shot transfer, the error remains low at 3.5 mm. The addition of a small amount of new samples for fine-tuning further improves accuracy. With 300 additional samples, the position RMSE reaches to approximately 1 mm.

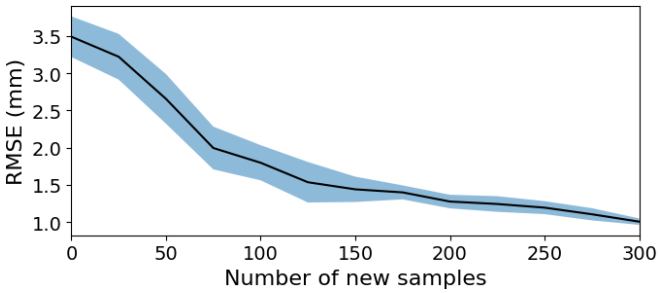

Fig. 2: Position estimation error with regards to the number of real images from the test sensor used to fine-tune model f_θ. Results with zero new tactile images are the zero-shot transfer errors.

REFERENCES

[1] J. Xu, S. Kim, T. Chen, A. R. Garcia, P. Agrawal, W. Matusik, and S. Sueda, "Efficient tactile simulation with differentiability for robotic manipulation," in *Conf. on Robot Learning*, 2023, pp. 1488–1498.

[2] A. Church, J. Lloyd, N. F. Lepora *et al.*, "Tactile sim-to-real policy transfer via real-to-sim image translation," in *Conference on Robot Learning*. PMLR, 2022, pp. 1645–1654.

[3] S. Suresh, Z. Si, S. Anderson, M. Kaess, and M. Mukadam, "Midastouch: Monte-carlo inference over distributions across sliding touch," in *Conference on Robot Learning*. PMLR, 2023, pp. 319–331.

[4] W. Yuan, S. Dong, and E. H. Adelson, "Gelsight: High-resolution robot tactile sensors for estimating geometry and force," *Sensors*, vol. 17, no. 12, p. 2762, 2017.

[5] M. Lambeta, P.-W. Chou, S. Tian, B. Yang, B. Maloon, V. R. Most, D. Stroud, R. Santos, A. Byagowi, G. Kammerer *et al.*, "Digit: A novel design for a low-cost compact high-resolution tactile sensor with application to in-hand manipulation," *IEEE Robotics and Automation Letters*, vol. 5, no. 3, pp. 3838–3845, 2020.

[6] I. H. Taylor, S. Dong, and A. Rodriguez, "Gelslim 3.0: High-resolution measurement of shape, force and slip in a compact tactile-sensing finger," in *IEEE Int. Conf. on Rob. & Auto.*, 2022, pp. 10 781–10 787.

[7] B. Ward-Cherrier, N. Pestell, L. Cramphorn, B. Winstone, M. E. Giannaccini, J. Rossiter, and N. F. Lepora, "The tactip family: Soft optical tactile sensors with 3d-printed biomimetic morphologies," *Soft robotics*, vol. 5, no. 2, pp. 216–227, 2018.

[8] A. Padmanabha, F. Ebert, S. Tian, R. Calandra, C. Finn, and S. Levine, "Omnitact: A multi-directional high-resolution touch sensor," in *IEEE Int. Conf. on Rob. & Auto.*, 2020, pp. 618–624.

[9] B. Romero, F. Veiga, and E. Adelson, "Soft, round, high resolution tactile fingertip sensors for dexterous robotic manipulation," in *IEEE Int. Conf. on Rob. & Auto.*, 2020, pp. 4796–4802.

[10] H. Sun, K. J. Kuchenbecker, and G. Martius, "A soft thumb-sized vision-based sensor with accurate all-round force perception," *Nature Machine Intelligence*, vol. 4, no. 2, pp. 135–145, 2022.

[11] S. Dong, D. K. Jha, D. Romeres, S. Kim, D. Nikovski, and A. Rodriguez, "Tactile-rl for insertion: Generalization to objects of unknown geometry," in *IEEE Int. Conf. on Rob. & Auto.*, 2021, pp. 6437–6443.

[12] S. Wang, M. Lambeta, P.-W. Chou, and R. Calandra, "Tacto: A fast, flexible, and open-source simulator for high-resolution vision-based tactile sensors," *IEEE Rob. & Aut. Let.*, vol. 7, pp. 3930–3937, 2022.

[13] Z. Si and W. Yuan, "Taxim: An example-based simulation model for gelsight tactile sensors," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 2361–2368, 2022.

[14] K. Patel, S. Iba, and N. Jamali, "Deep tactile experience: Estimating tactile sensor output from depth sensor data," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 9846–9853.

[15] Z. Si, Z. Zhu, A. Agarwal, S. Anderson, and W. Yuan, "Grasp stability prediction with sim-to-real transfer from tactile sensing," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2022, pp. 7809–7816.

[16] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *IEEE international conference on computer vision*, 2017, pp. 2223–2232.

[17] D. Ho, K. Rao, Z. Xu, E. Jang, M. Khansari, and Y. Bai, "Retinagan: An object-aware approach to sim-to-real transfer," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2021, pp. 10 920–10 926.