3D deformable surface reconstruction from visual and tactile input with geometric prior

Ioan Laurentiu Popa
Analog Devices Inc.
Cluj-Napoca,
Romania
Laurentiu.Popa@analog.com

Tudor Brezae, Paul Sucala, Robert Konievic, Levente Tamas
Automation Department
Technical University of Cluj-Napoca,
Romania
Levente.Tamas@aut.utcluj.ro

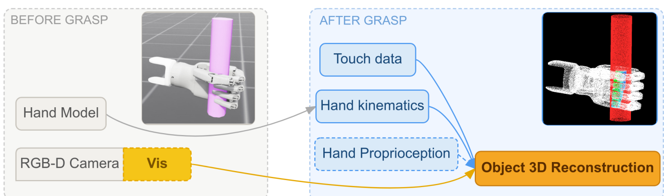


Fig. 1: The proposed architecture for the visual and tactile input for deformable surface reconstruction.

I. Introduction

Object grasping has once again become a central topic as increasingly affordable humanoid robots are emerging in both research and industrial settings. Beside the tactile sensors available in robotic arms, the visual data feed is used as well, motivating visual-tactile fusion for 3D surface estimation [7].

The combination of visual and tactile data as a multimodal input has already been introduced in several studies, such as [10], [14], [15]. Pioneering work on complementary visual-tactile input appear in [14] with a vision based global context and a touch based local structure proposal. A simulation-only based variant for visual and tactile data is proposed in [4] for rigid objects. A recent overview of the visual and tactile mixed data modalities is presented in [16].

In this paper we reconstruct 3D deformable surfaces of an object during the grasping by fusing the geometric prior information obtained from the visual data and the 3D shape information from the tactile sensing as this is shown in Fig. 1. The geometric prior is obtained from the depth camera using sampling consensus for predefined object classes such as cylindrical or spherical object of interest such as bottle or ball in a table-top setup. Real-robot experiments were conducted using the Inspire robotic hand, while simulated experiments were performed in IsaacSim using the same depth camera model and robot arm configuration. The work most closely related to ours is [14], which we extend by adding geometric primitive estimation at the visual sensing stage.

Supported by the Romanian National Authority for Scientific Research, project nr. PN-IV-P7-7.1-PTE-2024-0105 and by Analog Devices International Romania

II. Proposed approach

We propose a geometric prior based deformable object grasping using a humanoid arm both in simulation and real experiments. Our main goal is to estimate the degree of deformation using a geometric prior from camera data fused with the grasping information from the 6 DoF hand inverse kinematics (IK) and touch sensors as suggested in Figure 1.

A. Deformation and Contact Modelling in Simulation

For contact-aware deformable object manipulation, IsaacSim includes a physics-based Contact Sensor built on the PhysX Contact Report API [1], [5]. In our setup, to keep reliable tactile-like feedback we use a rigid cylinder and sphere with compliant contacts defined in the physics material. This choice aligns with prior IsaacSim tactile module [3], [6] that models soft interaction while keeping bodies rigid similar to TacSL which uses penalty-based soft-contact constraints to allow controlled interpenetration [7]. The manipulated objects in our case are a rigid cylinder and sphere with compliant-contact material parameters, which emulates deformable interaction. In our implementation, we introduce compliance by setting non-zero compliant-contact stiffness and damping in the object's material [2].

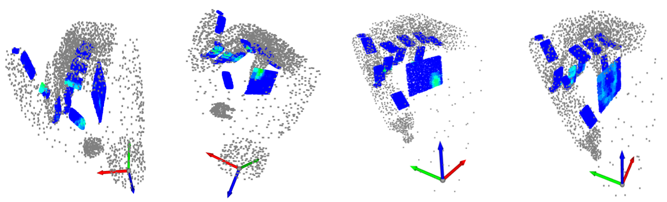
B. Anthropomorphic hand experimental setup

The real robot experiments used a 6-DoF anthropomorphic Inspire robot Hand, equipped with embedded tactile sensors across all five fingers and the palm. The hand is controlled via ROS2 through a Modbus TCP interface and provides both motor-current-based force feedback and high-resolution tactile pressure readings from distributed sensor arrays. The tactile data was reprojected in 3D using the forward kinematic (FK) of the hand, thus obtaining a point cloud which has 5 fields per point: position (x, y, z), RGB color, and 12-bit intensity data as this is visible in Figure 2.

We collected a multi-modal grasping dataset comprising more than 6,000 synchronized samples, containing camera RGB images, tactile heat-maps, 3D tactile point clouds with intensity, hand actuator states, and joint angles. The dataset is classified into deformable objects and non-deformable objects based on object compliance. Figure 3 shows sample data, presenting visual and tactile modalities during grasping.

Method		Cylinders								Sphere			
		250ml	330ml slim	330ml	500ml	500ml bot.	1L bot.	1.5L bot.	**Overall**	Tennis	White	Orange	**Overall**
	GT (mm)	24.0	29.0	33.0	33.0	32.5	39.5	46.0	-	32.0	38.5	29.4	-
PointNet	Rigid MAE	**0.6**	2.2	**0.2**	**0.1**	1.1	**0.6**	**0.6**	**0.7**	**1.0**	**5.9**	-	3.6
	Rigid Std	1.1	2.3	0.2	0.2	1.2	0.9	0.3	-	1.3	1.7	-	-
	Deform MAE	6.5	5.1	3.8	6.6	3.3	7.7	11.4	6.6	2.4	-	**3.4**	**3.1**
	Deform Std	5.1	2.0	2.5	3.6	3.7	3.0	2.6	-	3.1	-	2.7	-
SAC	Rigid MAE	5.3	3.2	2.9	6.8	4.9	4.6	10.7	5.4	-	-	-	-
	Rigid Std	6.8	3.9	4.2	5.6	5.1	3.8	3.6	-	-	-	-	-
	Deform MAE	3.7	8.0	6.5	7.4	4.3	13.4	14.0	8.1	-	-	-	-
	Deform Std	4.6	1.3	4.7	4.4	4.1	4.9	2.8	-	-	-	-	-

TABLE I: Rigid objects achieve sub-mm accuracy (0.7mm MAE), deformable objects show higher error (6.6mm MAE).



(a) (b) (c) (d)

Fig. 2: 3D tactile point cloud visualization. (a) deformable bottle, (b) rigid bottle, (c) deformable ball, (d) rigid ball.

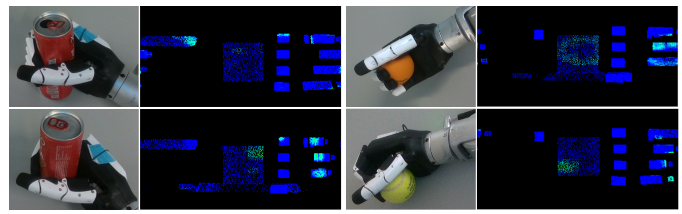


Fig. 3: RGB views with tactile heatmaps for cylinder (left) and sphere (right) grasping. Top: deformable, bottom: rigid.

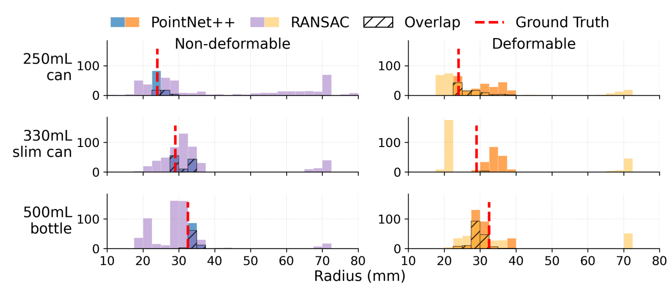
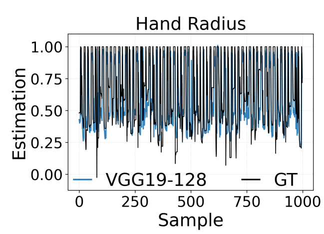
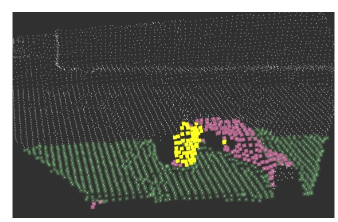


Fig. 4: Radius estimators comparison by object class. Cylinders of 24mm, 29mm and 32.5mm radius



(a) VGG radii estimate from RGB touch data



(b) RGB-D scene parsing (yellow cylinder, magenta hand)

Fig. 5: (a)VGG prediction on touch and (b)3D scene parsing

C. Learning from pointcloud data

We use PointNet++ [12] for hierarchical feature extraction from tactile point clouds. Unlike the original PointNet [11], which uses global pooling, it captures local geometric structures through set abstraction layers. Our regression approach is inspired by [9], which applies PointNet to 3D hand pose estimation. We employ data augmentation including z-axis rotation and ±10% scaling to improve generalization.

D. VGG19 for visual radii estimation

We evaluated a radius estimation pipeline for a 6-DOF hand in simulation. Fig. 5a shows a comparison between the VGG19 model [13] and the corresponding ground-truth values, with all values normalized to the range 0–1. The input consists of RGB-touch image data and associated values. The results indicate that the model effectively captures the underlying patterns in the data and produces accurate predictions, achieving a MAE of 0.6 for all cylinder types, and 0.04 for the tennis ball evaluation from the sphere category. We have also explored how transfer learning [8] techniques can be usedby training the VGG19 model on the left hand and then adapting it for the right hand and validating on it.

E. Deformable/non-deformable grasping sampling

Deformation detection uses proprioceptive and tactile feedback, encoding joint angles and contact forces into 2D images via cylindrical projection. A comparison between evaluated models and data configurations is summarized in Table I, while the distribution of the estimated radii is shown in Figure 4.

PointNet achieving sub-millimeter accuracy (MAE = 0.7mm) on rigid cylinders, with errors below 0.2mm on standard cans. For spheres, the cylinder-trained model generalizes well to rigid balls and deformable ones. SAC yields higher errors but requires no learning. VGG on simulated data achieves the lowest error (0.6mm cylinder, 0.04mm sphere), with real-world transfer showing 0.02mm on cylinders. Performance degrades on deformable objects due to compression during grasping. These results confirm the model learns surface curvature to radius mapping.

REFERENCES

[1] Isaac Lab DeformableObject: Finite-element deformable assets (source documentation). https://docs.robotsfan.com/isaaclab_official/v2.3.1/_modules/isaaclab/assets/deformable_object/deformable_object.html.

[2] Isaac Lab Physics Materials Config: Compliant contact stiffness/damping semantics. https://isaac-sim.github.io/IsaacLab/main/_modules/isaaclab/sim/spawners/materials/physics_materials_cfg.html.

[3] Isaac Lab physics materials config: Compliant contacts (implicit spring model). https://isaac-sim.github.io/IsaacLab/main/_modules/isaaclab/sim/spawners/materials/physics_materials_cfg.html.

[4] Isaac Lab sim.schemas: Contact reporting API can only be added to rigid bodies. https://isaac-sim.github.io/IsaacLab/v2.0.1/source/api/lab/isaaclab.sim.schemas.html.

[5] Isaac Lab sim.schemas: Deformable body properties and PhysX soft-body (FEA) description. https://isaac-sim.github.io/IsaacLab/v2.0.1/source/api/lab/isaaclab.sim.schemas.html.

[6] Isaac Sim Physics Fundamentals: Compliant contacts as spring–damper contact to approximate deformables with rigid bodies. https://docs.isaacsim.omniverse.nvidia.com/4.2.0/simulation_fundamentals.html.

[7] Iretiayo Akinola, Jie Xu, Jan Carius, Dieter Fox, and Yashraj Narang. TacSL: A Library for Visuotactile Sensor Simulation and Learning. *IEEE Transactions on Robotics*, 41:2645–2661, 2025.

[8] Gabriela Csurka. Domain Adaptation for Visual Applications: A Comprehensive Survey. *arXiv preprint arXiv:1702.05374*, 2017.

[9] Liuhao Ge, Zhou Ren, and Junsong Yuan. Point-to-Point Regression PointNet for 3D Hand Pose Estimation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018.

[10] Yangrong Liu, Jian Li, Huaiyu Wang, Ming Lu, Haorao Shen, and Qin Wang. TAPCNet: Tactile-Assisted Point Cloud Completion Network via Iterative Fusion Strategy. *IET Computer Vision*, 19(1):e70012, 2025.

[11] Charles R. Qi, Hao Su, Kaichun Mo, and Leonidas J. Guibas. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[12] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[13] Karen Simonyan and Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[14] Edward Smith, Roberto Calandra, Adriana Romero, Georgia Gkioxari, David Meger, Jitendra Malik, and Michal Drozdzal. 3D Shape Reconstruction from Vision and Touch. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 14193–14206. Curran Associates, Inc., 2020.

[15] Yuanbo Wang, Zhaoxuan Zhang, Jiajin Qiu, Dilong Sun, Zhengyu Meng, Xiaopeng Wei, and Xin Yang. Touch2Shape: Touch-Conditioned 3D Diffusion for Shape Exploration and Reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5656–5665, June 2025.

[16] Shixin Zhang, Zixi Chen, Yuan Gao, Weiwei Wan, Jianhua Shan, Hongxiang Xue, Fuchun Sun, Yiyong Yang, and Bin Fang. Hardware Technology of Vision-Based Tactile Sensor: A Review. *IEEE Sensors Journal*, 22(22):21410–21427, 2022.