DOFS: A Real-world 3D Deformable Object Dataset with Full Spatial Information for Dynamics Model Learning

Zhen Zhang, Xiangyu Chu, Yunxi Tang, K. W. Samuel Au
Department of Mechanical and Automation Engineering
The Chinese University of Hong Kong, HKSAR
Multi-scale Medical Robotics Center
zhzhen@link.cuhk.edu.hk

Abstract: This work proposes DOFS, a pilot dataset of 3D deformable objects (DOs) (e.g., elasto-plastic objects) with full spatial information (i.e., top, side, and bottom information) using a novel and low-cost data collection platform with a transparent operating plane. The dataset consists of active manipulation action, multi-view RGB-D images, well-registered point clouds, 3D deformed mesh, and 3D occupancy with semantics, using a pinching strategy with a two-parallel-finger gripper. In addition, we trained a neural network with the down-sampled 3D occupancy and action as input to model the dynamics of an elasto-plastic object. Our dataset and all CADs of the data collection system will be released soon on our website.

Keywords: Deformable Object Manipulation, Data Collection, State Representation

1 Introduction

Robot manipulation of 3D Deformable Objects is essential for many activities and applications in the real world, such as household [1, 2] and healthcare [3], and is still an open challenge despite extensive studies. Recently, data-driven solutions have shown impressive and promising results in 3D deformable object manipulation by learning-based approaches [4, 5], where sufficient data is essential to improve model training or policy learning. To obtain training data, some previous works collected synthetic data from simulators [6]. Still, there is an unavoidable gap between the real world and the simulator since the existing simulators cannot accurately simulate all real-world physical characteristics (e.g., friction, impact, and stiffness) [7]. To mitigate the gap, some researchers [8, 9, 10, 11] collect Real-World Data (RWD); for example, [8, 9] collects RGB-D images and point clouds, [10] collects 3D mesh models, [11] uses a professional system with 106 cameras to obtain the 3D reconstructions of deformed mesh. However, there are still some limitations. The data of [8, 9, 10, 11] are not full-spatial without considering the bottom side of 3D DOs as they are placed on the operating plane, which may lead to the dramatic dynamics difference caused by the hollow bottom as shown in Fig. 1. Besides, [11] is highly costly and complicated to replicate.

We propose a low-cost and novel data collection platform to collect the dataset of 3D elasto-plastic object manipulation with full spatial information. Our DOFS dataset includes 3D reconstructions of deformed meshes, corresponding actions, well-registered 3D point clouds, and multi-view RGB-D images. In addition, we also provide the 3D occupancy of the manipulation scenario, which is a very dense, high-resolution, and semantic-rich representation.
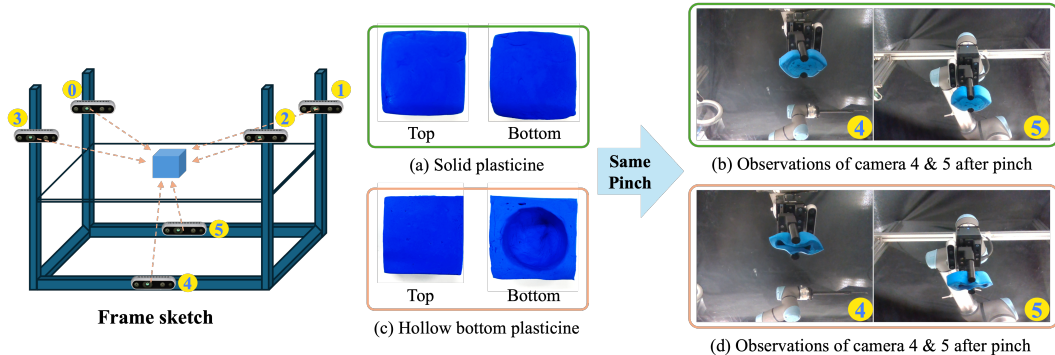
Figure 1: **Dynamics difference caused by the hollow bottom**. For two pieces of plasticine of the same appearance and size, one is solid and another is hollow, the deformation results are completely different with the same action.

2 Methods

To build our DOFS dataset, we first built a novel data collection platform including an aluminum frame with a transparent operating plane, a robotic arm, and 6 RGB-D cameras. We proposed a generalizable pipeline to process collected raw data and generate the structured dataset. Finally, we trained a dynamics model of the plasticine using our DOFS dataset to validate the quality of the dataset.

2.1 Hardware Setup

In our data collection system, we propose a novel hardware setup to capture the multi-view RGB-D images from the top, sides, and bottom of objects for full spatial information collection. To be specific, we used aluminum profiles to build a frame and installed a transparent acrylic board in the middle of the frame as the operating plane, where the objects are placed. We use 6 Intel RealSense D435i cameras and install 4 cameras on top of it and 2 cameras on the bottom from different positions and angles respectively to collect multi-view RGB-D images of the manipulation scenario to obtain dense point clouds, as shown in Fig. 2.
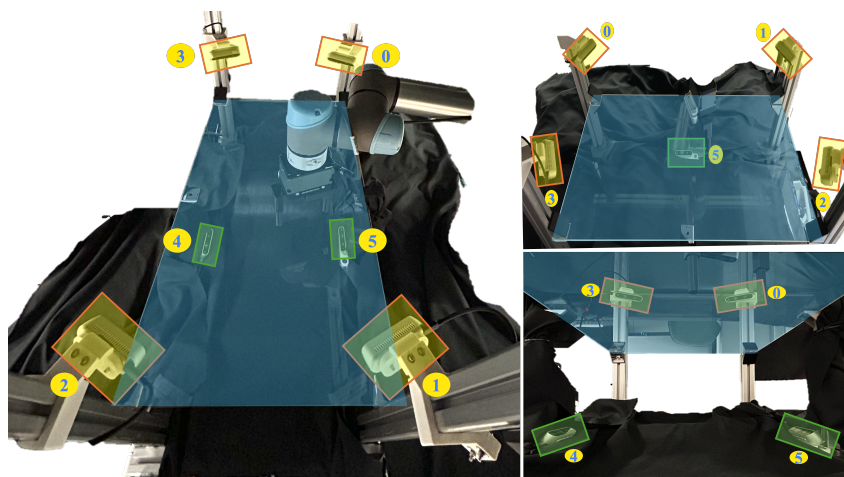

Figure 2: **Hardware setup of our data collection platform**. The blue part is a transparent acrylic board that serves as an operating plane. The orange parts are 4 cameras above the plane to collect top and side information. The green parts are 2 cameras below the plane to collect bottom-side information.
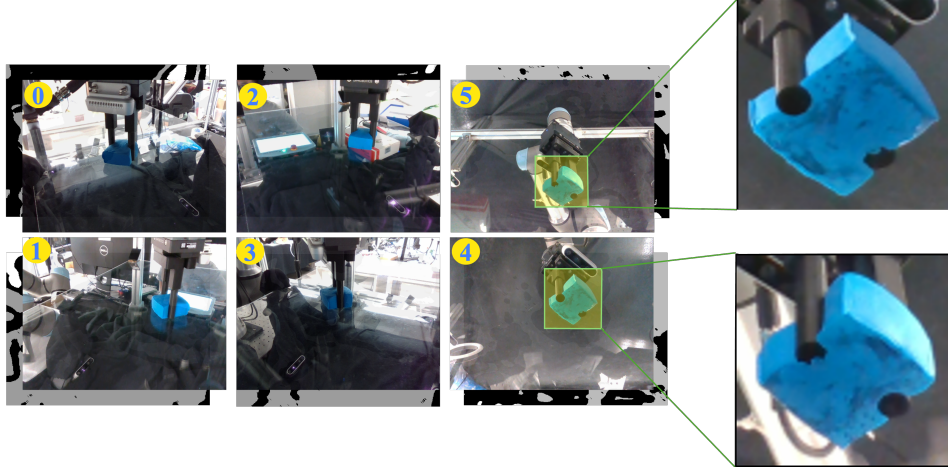
Figure 3: **Six RGB-D images captured from RealSense D435i**. Four images from the cameras installed above the operating plane show the top and side information of the plasticine. Two images from the cameras installed below the operating plane show the bottom side information of the plasticine.

2.2 Data Collection and Processing

We use ROS [12] to synchronize and record the actions of the robotic arm and 6 RGB-D cameras during manipulation. We use Iterative Closest Point (ICP) [13] to register the point cloud generated from 6 RGB-D cameras to obtain the full spatial information. After that, we used geometry information and registered point clouds to reconstruct the deformed meshes. Moreover, we selected the voxel size $[0.002, 0.002, 0.002]m$ to generate the 3D occupancy data and assign the semantic label for the operating plane, two fingers, and deformable objects.

3 Results

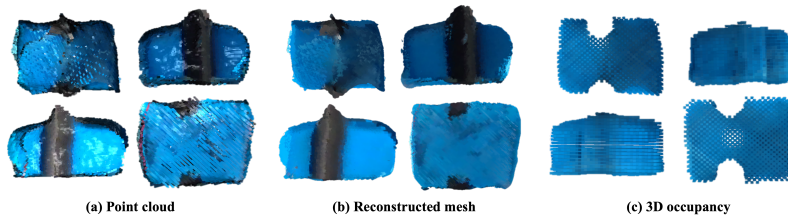(a) Point cloud (b) Reconstructed mesh (c) 3D occupancy

Figure 4: **Data visualization of one frame during manipulation.** (a): Well-registered point cloud of pinched plasticine without background. (b): Reconstructed deformed mesh. (c): 3D occupancy of plasticine.

We use 6 well-calibrated cameras to collect 6 RGB-D images shown in Fig. 3 and record the action (i.e., the 3D positions of two fingers) simultaneously during manipulation. We use 6 RGB-D images and their intrinsic and extrinsic matrix to generate dense point clouds shown in Fig. 4 (a) for deformed mesh reconstruction shown in Fig. 4 (b). Specifically, We use UR5e with a two-finger paralleled gripper of DH Robotics PGI-140-80 to manipulate the plasticine. We discrete one pinch action into 30 steps to pinch deformable objects step by step. Our current DOFS dataset contains 150 pinches(i.e., 4500 frames), and one frame captures the object deformation shown in Figure 4 (c). In addition, each frame contains the following information:

- 3D deformed mesh of the plasticine

- 3D positions of two fingers
- Multi-view RGB-D images from 6 cameras
- Registered 3D point clouds
- 3D occupancy with semantics

We embedded a cartoon model into the bottom of the plasticine and took it out, shown in Fig. 5 (a). We capture the data of the plasticine placed on the operating plane and visualize the point cloud and deformed mesh with geometry details shown in Fig. 5 (b) and Fig. 5 (c), respectively, for reporting full-spatial results.
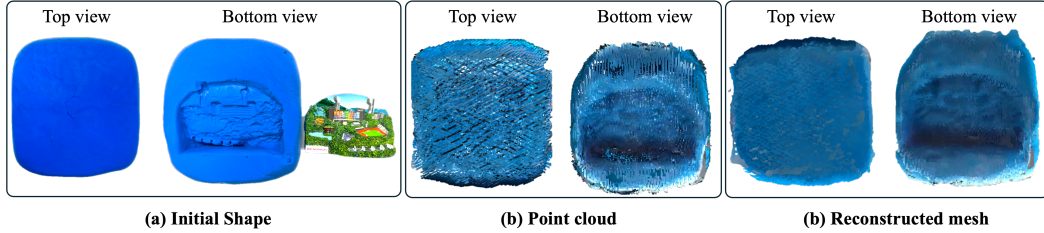

(a) Initial Shape   **(b) Point cloud**   **(b) Reconstructed mesh**

Figure 5: **Visualization of Full spatial information collection**. (a): We embed the cartoon model into the bottom of the plasticine and take it out. (b): The visualization of the point cloud of plasticine. (c): The deformed mesh of plasticine.

To evaluate the quality of our DOFS dataset, we downsample the generated 3D occupancy and use it to train a dynamics model of the plasticine. We trained it for 8 epochs using a machine with Intel 12700K x 12 Core Processor CPU and NVIDIA GeForce RTX 4080 GPU (16GB memory). The preliminary experiment shown in Fig. 6 shows that our model can predict the state of plasticine after an action is applied.
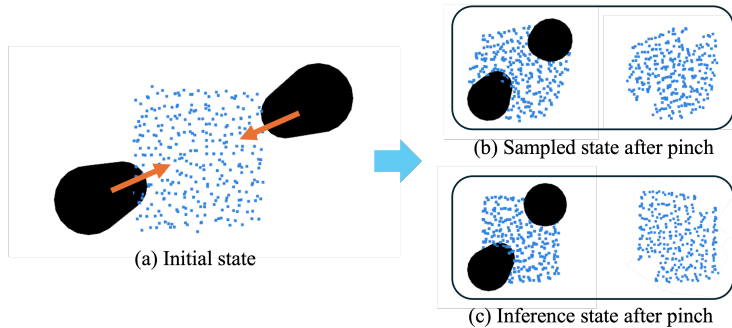

(a) Initial state   (b) Sampled state after pinch   (c) Inference state after pinch

Figure 6: **State transformation after pinching the plasticine**. (a): Visualization of the initial state of the plasticine and two fingers, the yellow arrow indicates the pinch direction. (b): The ground-truth state after the pinch. (c): The inference state of the learned dynamics model.

4   Conclusion

Our low-cost and novel data collection platform can effectively generate detailed 3D deformed meshes and dense, high-resolution 3D occupancy data. The learned dynamics model shows great potential for deformable object manipulation. To promote our DOFS dataset and system, we will choose 3D DOs with different physical properties (e.g., stiffness) and geometry shapes to diversify our dataset. In addition, we will change the gripper shape to apply other manipulation strategies (e.g., cut, roll). We consider that our DOFS dataset and data collection system have the potential to advance the research on robotic fine and dexterous manipulation, and we look forward to sharing this DOFS dataset and platform available with the community.

References

[1] R. Hoque, K. Shivakumar, S. Aeron, G. Deza, A. Ganapathi, A. Wong, J. Lee, A. Zeng, V. Vanhoucke, and K. Goldberg. Learning to fold real garments with one arm: A case study in cloud-based robotics research, 2022. URL https://arxiv.org/abs/2204.10297.

[2] Y. Avigal, L. Berscheid, T. Asfour, T. Kröger, and K. Goldberg. Speedfolding: Learning efficient bimanual folding of garments, 2022. URL https://arxiv.org/abs/2208.10552.

[3] P. M. Scheikl, E. Tagliabue, B. Gyenes, M. Wagner, D. Dall'Alba, P. Fiorini, and F. Mathis-Ullrich. Sim-to-real transfer for visual reinforcement learning of deformable object manipulation for robot-assisted surgery. *IEEE Robotics and Automation Letters*, 8(2):560–567, Feb. 2023. ISSN 2377-3774. doi:10.1109/lra.2022.3227873. URL http://dx.doi.org/10.1109/LRA.2022.3227873.

[4] W. Yan, A. Vangipuram, P. Abbeel, and L. Pinto. Learning predictive representations for deformable objects using contrastive estimation, 2020. URL https://arxiv.org/abs/2003.05436.

[5] H. Shi, H. Xu, S. Clarke, Y. Li, and J. Wu. Robocook: Long-horizon elasto-plastic object manipulation with diverse tools, 2023. URL https://arxiv.org/abs/2306.14447.

[6] I. Huang, Y. Narang, C. Eppner, B. Sundaralingam, M. Macklin, R. Bajcsy, T. Hermans, and D. Fox. Defgraspsim: Physics-based simulation of grasp outcomes for 3d deformable objects. *IEEE Robotics and Automation Letters*, 7(3):6274–6281, 2022. doi:10.1109/LRA.2022.3158725.

[7] S. Höfer, K. Bekris, A. Handa, J. C. Gamboa, M. Mozifian, F. Golemo, C. Atkeson, D. Fox, K. Goldberg, J. Leonard, C. Karen Liu, J. Peters, S. Song, P. Welinder, and M. White. Sim2real in robotics and automation: Applications and challenges. *IEEE Transactions on Automation Science and Engineering*, 18(2):398–400, 2021. doi:10.1109/TASE.2021.3064065.

[8] A. Bartsch, C. Avra, and A. B. Farimani. Sculptbot: Pre-trained models for 3d deformable object manipulation, 2023. URL https://arxiv.org/abs/2309.08728.

[9] C. Li, Z. Ai, T. Wu, X. Li, W. Ding, and H. Xu. Deformnet: Latent space modeling and dynamics prediction for deformable object manipulation, 2024. URL https://arxiv.org/abs/2402.07648.

[10] H. Shi, H. Xu, Z. Huang, Y. Li, and J. Wu. Robocraft: Learning to see, simulate, and shape elasto-plastic objects with graph networks, 2022. URL https://arxiv.org/abs/2205.02909.

[11] J. Obrist, M. Zamora, H. Zheng, J. Zarate, R. K. Katzschmann, and S. Coros. Pokeflex: Towards a real-world dataset of deformable objects for robotic manipulation, 2024. URL https://arxiv.org/abs/2409.17124.

[12] M. Quigley. Ros: an open-source robot operating system. In *IEEE International Conference on Robotics and Automation*, 2009. URL https://api.semanticscholar.org/CorpusID:6324125.

[13] J. Park, Q.-Y. Zhou, and V. Koltun. Colored point cloud registration revisited. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 143–152, 2017. doi:10.1109/ICCV.2017.25.