Towards Versatile Manipulation of Deformable Objects

Alberta Longhini
KTH Royal Institute of Technology, albertal@kth.se

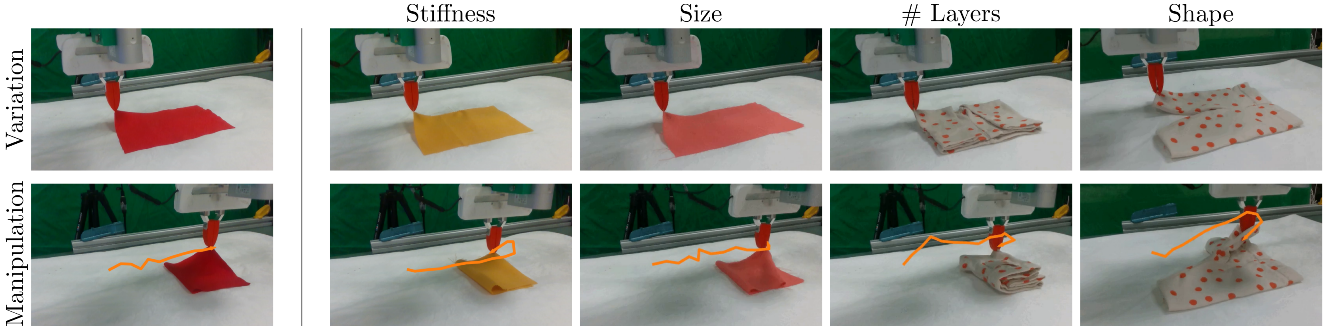

Fig. 1: **Versatile manipulation of cloth-like deformable objects:** Our work contributes benchmarking, modeling, perception, and control techniques for deformable objects, which together enable the robot to dynamically adapt the manipulation trajectory (in orange) to unseen variations in real-world object properties, enhancing generalization across diverse environments.

I. Research Problem

In spite of the rapid advancements in AI [31, 1], tasks like laundry, tidying, and general household assistance remain challenging for robots to scale due to their limited ability to generalize manipulation skills across diverse environments [43]. Fundamental for these tasks is the manipulation of deformable objects, which is particularly difficult due to their non-linear dynamics and infinite degrees of freedom [39]. Yet, improving robotic manipulation of deformables has far-reaching benefits, from assisting aging populations to enhancing efficiency and sustainability in industries such as healthcare, manufacturing, food preparation, and logistics [12, 9].

Current methods for manipulating deformable objects are largely dominated by either open-loop planning pipelines or end-to-end data-driven approaches [27]. These methods tend to overfit to specific training distributions and object instances, and they typically require large datasets due to the high-dimensional state space of deformables. As a result, these approaches lack the flexibility to adapt online, especially when faced with unseen variations. We argue that the missing piece is a sample-efficient methodology to effectively close the action-perception loop, enabling real-time sensory feedback and adaptive control. While recent advances in generative AI and imitation learning have shown improvements in cloth manipulation [42, 4], these solutions rely on costly human demonstrations and still lack generalization across variations in object properties and environments.

This research seeks to enhance the *versatility* of robotic manipulation for deformable objects, that is, in this context, the ability to generalize across diverse objects and environmental conditions [4]. We propose a structured, model-based approach to manipulation that explicitly leverages feedback for online adaptation. Cloth-like deformable objects, such as textiles, will serve as testbeds for developing new methodologies.

II. Contributions

Achieving *versatility* in manipulation requires robots to perceive, interpret, and adapt to dynamic environments. This, in turn, demands a comprehensive characterization and understanding of deformable object properties and dynamics, alongside advanced perception tools for real-time feedback. Integrating these elements within a closed action-perception loop enables robots to refine their strategies, ensuring adaptive and robust deformable object manipulation.

Characterization: Effective characterization of textile properties is essential for benchmarking tasks like folding and dressing, yet existing approaches lack a standardized taxonomy to thoroughly assess generalization. Most prior work categorizes objects (e.g., t-shirts, pants) [22, 44, 11] or considers material properties [34], but overlooks the role of textile construction techniques in defining physical behavior during manipulation. In our work [23], we introduced a taxonomy that integrates fiber material and construction techniques, providing a more comprehensive framework for textile characterization. However, precise mechanical property measurements remain a challenge due to the lack of accessible, non-destructive tools [16, 8]. To address this, in a follow-up work [10], we contributed a measurement framework grounded in textile industry standards to facilitate accurate property assessments and more robust benchmarking for cloth manipulation.

Modeling and Perception: A key challenge in deformable object manipulation is modeling deformation dynamics, where existing approaches fall into two categories: physics-based models, which allow long-horizon simulations but suffer from the sim-to-real gap [5], and data-driven models, which offer flexible state representations but are prone to accumulation errors [39]. Most prior work on data-driven models assumes fixed textile properties [30, 21, 14, 36, 7], limiting generalization across object variations. Our work tackled this limitation by integrating interactive perception with data-driven dynamics models, addressing two key challenges: property identification (inferring elasticity, friction, and stiffness) and state estimation (providing real-time feedback for adaptation).

For property identification, in [25], we leveraged textile industry knowledge to encode force measurements as indicators of elasticity, improving model generalization. However, this method required a closed-form solution for each property, limiting scalability. To overcome this, in [24], we introduced a self-supervised adaptation module that learned a latent representation of textile properties, allowing the dynamics model to adjust predictions based on observed object behavior. The proposed perception-driven adaptation significantly enhanced the generalization of model predictions across diverse textiles.

A key limitation of the data-driven models we proposed was their open-loop formulation due to the challenges in state estimation, preventing them from updating state representations as new observations become available. Closing this loop is crucial for adaptive manipulation, as models can be inaccurate and accumulate errors over time. However, real-time state estimation remains challenging due to occlusions and continuous deformations [15, 6, 19]. To address this, we developed two complementary approaches. The first, presented in [26], introduced a graph-based tracking method that refines 3D state estimations from RGB images using Gaussian Splatting (GS) [17], enhancing accuracy in complex deformations and self-occluded configurations. The second, in [28], shifted from dense tracking to task-specific semantic point cloud representations, leveraging geometric priors to resolve ambiguities in crumpled configurations. By improving state estimation, these methods contributed to closing the action-perception loop, enabling model correction and replanning.

Control: While early work highlighted the importance of optimizing deformable object manipulation trajectories [20], much of the current research still relies on open-loop pick-and-place planning, disregarding relevant information within the pick-and-place trajectory [38, 32, 21, 33, 41]. This is primarily due to the difficulty of simultaneously estimating and tracking cloth states during manipulation. The modeling and perception advances discussed earlier provide a foundation for addressing this limitation, enabling the closure of the action-perception loop and facilitating adaptive manipulation. In [28], we integrated the aforementioned state estimation and modeling components into a model-based manipulation framework, proposing a closed-loop approach to optimize cloth folding trajectories. Unlike previous work, our method leverages continuous feedback to dynamically adapt to changes in object size, shape, and physical properties, enabling more versatile manipulation. We evaluated the method's generalization across a diverse set of real-world textiles, characterized using the taxonomy and measurement tools introduced in [23, 10], demonstrating improved adaptability to textile variations. The results, shown in Figure 1, underscore the effectiveness of closing the action-perception loop in achieving versatile, real-world deformable object manipulation.

III. Future Work

The proposed methods, along with their limitations, open several promising research directions, including improving sample efficiency through hybrid model-based and model-free approaches [13] and extending the framework to a broader range of deformable objects such as cables, dough, and food. This section will focus on two particularly compelling future directions that build upon the foundations of this work, pushing the boundaries of *versatile* deformable object manipulation and opening new avenues for investigation.

Closing the sim-to-real gap: Simulating object dynamics is essential for training robotic policies without relying on extensive real-world interactions. However, existing physics-based models [3] face analytical and computational limitations, especially for deformable objects, resulting in a persistent sim-to-real gap. In contrast, generative data-driven world models offer the potential to learn complex dynamics directly from the real world [37, 29], but currently lack physical fidelity [2]. To overcome the limitations of both approaches, we propose hybrid dynamics models that combine physics-based structure with generative residuals to adapt from real-world observations and close the sim-to-real gap. Building on our prior work, we aim to extend our adaptation module to refine these residuals online using real-world feedback. Yet, closing the sim-to-real gap via hybrid world models remains challenging due to the lack of a shared state space between vision-based generative models and physics-based simulations. We suggest using GS as a bridge between 3D object states and visual observations, as proposed in [26]. This direction will enable adaptive, data-efficient learning while ensuring robust physical reasoning for robotic manipulation tasks with complex dynamics.

Broadening semantic understanding: The discussion thus far primarily focused on the physical and mechanical properties of textiles that influence cloth manipulation. However, object properties exist at multiple levels of abstraction, from measurable physical attributes (e.g., elasticity) to human semantic descriptors (e.g., stretchy or soft). Bridging these levels requires robots to ground semantic knowledge in the physical world through multisensory perception, integrating vision, haptic, and tactile inputs [40]. Future work will focus on developing a framework that combines these sensory modalities with the reasoning capabilities of vision-language models (VLMs) [18, 35] to ground semantic properties in robot sensing and enable robots to infer physical and semantic attributes. By leveraging interactive perception, robots will actively explore objects through actions like pulling or poking, refining their physical and semantic understanding of object properties. This direction not only promises to enhance the understanding of deformable objects for robotic systems but also establishes a foundation for richer language understanding and more *versatile* instruction formats, advancing both robotic autonomy and human-robot collaboration.

References

[1] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.

[2] Niket Agarwal et al. Cosmos world foundation model platform for physical ai. *arXiv preprint arXiv:2501.03575*, 2025.

[3] Genesis Authors. Genesis: A universal and generative physics engine for robotics and beyond, December 2024. URL https://github.com/Genesis-Embodied-AI/Genesis.

[4] Kevin Black, Noah Brown, Danny Driess, Adnan Esmail, Michael Equi, Chelsea Finn, Niccolo Fusai, Lachy Groom, Karol Hausman, Brian Ichter, et al. π_0: A vision-language-action flow model for general robot control. *arXiv preprint arXiv:2410.24164*, 2024.

[5] David Blanco-Mulero, Oriol Barbany, Gokhan Alcan, Adrià Colomé, Carme Torras, and Ville Kyrki. Benchmarking the sim-to-real gap in cloth manipulation. *IEEE Robotics and Automation Letters*, 2024.

[6] Cheng Chi and Shuran Song. Garmentnets: Category-level pose estimation for garments via canonical space shape completion. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3324–3333, 2021.

[7] Yuhong Deng, Chongkun Xia, Xueqian Wang, and Lipeng Chen. Deep reinforcement learning based on local gnn for goal-conditioned deformable object rearranging. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1131–1138. IEEE, 2022.

[8] Peirce F. T. The “handle” of cloth as a measurable quality. *Journal of the textile institute*, 21:T377–T416, 1930.

[9] Teresa Fischer. Textile turning point: Robotics for the apparel industry, 2023.

[10] Irene Garcia-Camacho, Alberta Longhini, Michael Welle, Guillem Alenyà, Danica Kragic, and Júlia Borràs. Standardization of cloth objects and its relevance in robotic manipulation. *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024.

[11] Yuying Ge, Ruimao Zhang, Xiaogang Wang, Xiaoou Tang, and Ping Luo. Deepfashion2: A versatile benchmark for detection, pose estimation, segmentation and re-identification of clothing images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5337–5345, 2019.

[12] Cristina Getson and Goldie Nejat. Socially assistive robots helping older adults through the pandemic and life after covid-19. *Robotics*, 10(3):106, 2021.

[13] Nicklas Hansen, Hao Su, and Xiaolong Wang. Td-mpc2: Scalable, robust world models for continuous control. *arXiv preprint arXiv:2310.16828*, 2023.

[14] Zixuan Huang, Xingyu Lin, and David Held. Mesh-based dynamics with occlusion reasoning for cloth manipulation. *arXiv preprint arXiv:2206.02881*, 2022.

[15] Zixuan Huang, Xingyu Lin, and David Held. Self-supervised cloth reconstruction via action-conditioned cloth tracking. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 7111–7118. IEEE, 2023.

[16] Sueo Kawabata and Masako Niwa. Fabric performance in clothing and clothing manufacture. *Journal of the Textile Institute*, 80(1):19–50, 1989.

[17] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4):139–1, 2023.

[18] Moo Jin Kim, Karl Pertsch, Siddharth Karamcheti, Ted Xiao, Ashwin Balakrishna, Suraj Nair, Rafael Rafailov, Ethan Foster, Grace Lam, Pannag Sanketi, et al. Openvla: An open-source vision-language-action model. *arXiv preprint arXiv:2406.09246*, 2024.

[19] Yinxiao Li, Yan Wang, Michael Case, Shih-Fu Chang, and Peter K Allen. Real-time pose estimation of deformable objects using a volumetric approach. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1046–1052. IEEE, 2014.

[20] Yinxiao Li, Yonghao Yue, Danfei Xu, Eitan Grinspun, and Peter K Allen. Folding deformable objects using predictive simulation and trajectory optimization. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 6000–6006. IEEE, 2015.

[21] Xingyu Lin, Yufei Wang, Zixuan Huang, and David Held. Learning visible connectivity dynamics for cloth smoothing. In *Conference on Robot Learning*, pages 256–266. PMLR, 2022.

[22] Ziwei Liu, Ping Luo, Shi Qiu, Xiaogang Wang, and Xiaoou Tang. Deepfashion: Powering robust clothes recognition and retrieval with rich annotations. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1096–1104, 2016.

[23] Alberta Longhini, Michael C Welle, Ioanna Mitsioni, and Danica Kragic. Textile taxonomy and classification using pulling and twisting. *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 7564–7571, 2021.

[24] Alberta Longhini, Marco Moletta, Alfredo Reichlin, Michael C Welle, David Held, Zackory Erickson, and Danica Kragic. Edo-net: Learning elastic properties of deformable objects from graph dynamics. *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3875–3881, 2023.

[25] Alberta Longhini, Marco Moletta, Alfredo Reichlin, Michael C Welle, Alexander Kravberg, Yufei Wang, David Held, Zackory Erickson, and Danica Kragic. Elastic context: Encoding elasticity for data-driven models of textiles. *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1764–1770, 2023.

[26] Alberta Longhini, Marcel Büsching, Bardienus P Duisterhof, Jens Lundell, Jeffrey Ichnowski, Mårten Björkman, and Danica Kragic. Cloth-splatting: 3d cloth state estimation from rgb supervision. *Conference on Robot Learning (CoRL)*, 2024.

[27] Alberta Longhini, Yufei Wang, Irene Garcia-Camacho, David Blanco-Mulero, Marco Moletta, Michael Welle, Guillem Alenyà, Hang Yin, Zackory Erickson, David Held, et al. Unfolding the literature: A review of robotic cloth manipulation. *Annual Review of Control, Robotics, and Autonomous Systems*, 8, 2024.

[28] Alberta Longhini, Michael C Welle, Zackory Erickson, and Danica Kragic. Adafold: Adapting folding trajectories of cloths via feedback-loop manipulation. *IEEE Robotics and Automation Letters (RA-L)*, 2024.

[29] Pietro Mazzaglia, Tim Verbelen, Bart Dhoedt, Aaron Courville, and Sai Rajeswar. Genrl: Multimodal-foundation world models for generalization in embodied agents. *arXiv preprint arXiv:2406.18043*, 2024.

[30] Tobias Pfaff, Meire Fortunato, Alvaro Sanchez-Gonzalez, and Peter W Battaglia. Learning mesh-based simulation with graph networks. *arXiv preprint arXiv:2010.03409*, 2020.

[31] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022.

[32] Daniel Seita, Aditya Ganapathi, Ryan Hoque, Minho Hwang, Edward Cen, Ajay Kumar Tanwani, Ashwin Balakrishna, Brijen Thananjeyan, Jeffrey Ichnowski, Nawid Jamali, et al. Deep imitation learning of sequential fabric smoothing from an algorithmic supervisor. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 9651–9658. IEEE, 2020.

[33] Haochen Shi, Huazhe Xu, Samuel Clarke, Yunzhu Li, and Jiajun Wu. Robocook: Long-horizon elasto-plastic object manipulation with diverse tools. *arXiv preprint arXiv:2306.14447*, 2023.

[34] Matti Strese, Lara Brudermueller, Jonas Kirsch, and E. Steinbach. Haptic material analysis and classification inspired by human exploratory procedures. *IEEE Transactions on Haptics*, 13:404–424, 2020.

[35] Yi Ru Wang, Jiafei Duan, Dieter Fox, and Siddhartha Srinivasa. Newton: Are large language models capable of physical reasoning? *arXiv preprint arXiv:2310.07018*, 2023.

[36] Yufei Wang, David Held, and Zackory Erickson. Visual haptic reasoning: Estimating contact forces by observing deformable object interactions. *IEEE Robotics and Automation Letters*, 7(4):11426–11433, 2022.

[37] Yufei Wang et al. Robogen: Towards unleashing infinite data for automated robot learning via generative simulation. *ICML*, 2024.

[38] Thomas Weng, Sujay Man Bajracharya, Yufei Wang, Khush Agrawal, and David Held. Fabricflownet: Bimanual cloth manipulation with a flow-based policy. In *Conference on Robot Learning*, pages 192–202. PMLR, 2022.

[39] Hang Yin, Anastasia Varava, and Danica Kragic. Modeling, learning, perception, and control methods for deformable object manipulation. *Science Robotics*, 6(54): eabd8803, 2021.

[40] Samson Yu, Kelvin Lin, Anxing Xiao, Jiafei Duan, and Harold Soh. Octopi: Object property reasoning with large tactile-language models. *arXiv preprint arXiv:2405.02794*, 2024.

[41] Kaifeng Zhang, Baoyu Li, Kris Hauser, and Yunzhu Li. Adaptigraph: Material-adaptive graph-based neural dynamics for robotic manipulation. *arXiv preprint arXiv:2407.07889*, 2024.

[42] Tony Z Zhao, Jonathan Tompson, Danny Driess, Pete Florence, Kamyar Ghasemipour, Chelsea Finn, and Ayzaan Wahid. Aloha unleashed: A simple recipe for robot dexterity. *arXiv preprint arXiv:2410.13126*, 2024.

[43] Jihong Zhu, Andrea Cherubini, Claire Dune, David Navarro-Alarcon, Farshid Alambeigi, Dmitry Berenson, Fanny Ficuciello, Kensuke Harada, Xiang Li, Jia Pan, et al. Challenges and outlook in robotic manipulation of deformable objects. *arXiv preprint arXiv:2105.01767*, 2021.

[44] Thomas Ziegler, Judith Butepage, Michael C Welle, Anastasiia Varava, Tonci Novkovic, and Danica Kragic. Fashion landmark detection and category classification for robotics. In *2020 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)*, pages 81–88. IEEE, 2020.