Frontiers in Robotics and Automation **Volume 3 Issue 1, 2026**
ISSN: 3079-644X

Closed-Loop Deformation Control of Soft Manipulators for Fruit Harvesting Fusing Tactile and Depth Vision

Robert J. Anderson, Sarah E. Mitchell, Thomas L. Hughes

Department of Plant and Environmental Sciences, University of Copenhagen, Frederiksberg 1871, Denmark

Abstract

The automation of agricultural harvesting presents a distinct set of challenges, primarily driven by the unstructured nature of the environment and the biological variability of the produce. Soft robotics has emerged as a promising avenue for handling delicate crops such as tomatoes, apples, and strawberries due to the inherent compliance of the materials used. However, the non-linear deformation characteristics of soft actuators make precise control difficult, often leading to either insufficient grasping force or damage to the fruit. This paper introduces a novel closed-loop deformation control framework that fuses depth vision with distributed tactile sensing to enhance the grasping efficacy of pneumatically actuated soft manipulators. By utilizing a depth camera to generate a preliminary volumetric estimation of the target and integrating real-time feedback from capacitive tactile sensors embedded within the gripper fingers, the system dynamically adjusts the internal pneumatic pressure to optimize the contact profile. We employ a sensor fusion algorithm that mitigates the limitations of individual modalities, such as visual occlusion and tactile sparsity. Experimental validation demonstrates that this multi-modal approach significantly reduces surface bruising on harvested fruits while maintaining a high success rate in detachment. The results indicate that fusing tactile and visual data allows for robust compensation of mechanical hysteresis and external disturbances, marking a substantial advancement in autonomous harvesting technologies.

Keywords

Soft Robotics, Sensor Fusion, Agricultural Automation, Tactile Sensing

1. Introduction

The global agricultural sector is currently facing a critical labor shortage, necessitating the rapid development of autonomous systems capable of performing labor-intensive tasks such as harvesting. Unlike industrial manufacturing, where environments are structured and objects are uniform, agriculture requires robots to interact with biological structures that vary significantly in size, shape, and stiffness. Traditional rigid robotic manipulators, while precise, often lack the adaptability required to handle delicate fruits without causing mechanical damage [1]. Consequently, the field has seen a paradigm shift toward soft robotics, where end-effectors are constructed from compliant materials like silicone elastomers, allowing them to passively conform to the geometry of the target object. Despite their advantages regarding passive compliance, soft manipulators introduce complex control problems. The materials used exhibit non-linear viscoelastic behavior, hysteresis, and infinite degrees of freedom, making it difficult to model their deformation using standard kinematic approaches [2]. In an open-loop control scenario, a soft gripper might apply excessive force to a smaller fruit or fail to secure a larger one, as the relationship between input pressure and resultant deformation is not constant across different contact scenarios. To address this, feedback mechanisms are essential. However, reliance on a single sensing modality often proves insufficient in the chaotic foliage of a crop row. Vision systems, particularly those utilizing depth sensing, provide crucial information regarding the global position and approximate geometry of the fruit. Yet, visual data is frequently compromised by occlusion from leaves, branches, or the manipulator itself during the grasping phase [3]. Conversely, tactile sensors provide direct measurements of contact forces and local deformation but lack information about the global context of the object. This paper proposes a unified control architecture that fuses these two distinct sensory inputs. By correlating the volumetric data from depth vision with the contact stress maps from tactile sensors, we establish a closed-loop controller capable of regulating the deformation of the soft fingers in real-time. This ensures that the grasping force remains within a safe margin to prevent bruising while being sufficient to detach the fruit from its peduncle.

2. Related Work

2.1 Soft Manipulation in Agriculture

The application of soft robotics in agriculture has expanded rapidly over the last decade. Early prototypes focused on pneumatic network actuators that bend upon inflation, mimicking the grasping motion of human fingers. These designs leveraged the natural compliance of the material to handle crops without complex control algorithms [4]. However, as the demand for higher harvesting speeds and lower damage rates increased, the limitations of purely passive compliance became apparent. Researchers began investigating variable stiffness actuators and granular jamming techniques to provide a wider range of grasping forces. Despite these hardware improvements, the lack of sensory feedback often resulted in suboptimal performance when the robot encountered fruits with varying ripeness and structural integrity [5].

2.2 Vision-Based Control Systems

Visual servoing is the dominant method for guiding robotic manipulators in unstructured environments. In the context of harvesting, RGB-D cameras are extensively used to detect and localize fruits. Advanced machine learning models, such as convolutional neural networks, have been employed to segment fruit from foliage and estimate the optimal approach vector [6]. While these systems are effective for approach planning, they struggle during the terminal phase of grasping. Once the end-effector comes into close proximity with the fruit, the camera view is often obstructed. Furthermore, depth sensors suffer from noise and lower resolution at close ranges, leading to inaccuracies in estimating the exact volume and surface curvature of the target [7]. This visual uncertainty necessitates a secondary feedback loop to verify successful acquisition and safe handling.

2.3 Tactile Sensing and Sensor Fusion

Tactile sensing in soft robotics is a developing field, often hindered by the difficulty of integrating rigid sensors into stretchable materials. Recent innovations involve liquid metal channels, optical fibers, and conductive fabrics that can withstand the deformation of the actuator [8]. Studies have shown that tactile feedback can detect slip events and estimate object stiffness, which is critical for grasping delicate items. However, tactile sensors only provide local data at the points of contact. Without knowledge of the overall object shape, tactile-driven control can lead to unstable grasps. The fusion of vision and touch has been explored in industrial bin-picking applications, where it has been shown to improve grasp stability significantly [9]. This paper extends those concepts into the agricultural domain, addressing the specific challenges of soft material hysteresis and biological variability in fruit harvesting.

ISSN: 3079-644X

3. System Architecture

3.1 Pneumatic Soft Manipulator Design

The end-effector utilized in this study is a four-fingered soft gripper fabricated from a platinum-cure silicone elastomer. Each finger contains a series of internal pneumatic chambers designed to induce a bending motion when pressurized. The geometry of the chambers is optimized to maximize the contact area with spherical or continuous objects, typical of fruits like tomatoes and apples. The actuation system comprises a programmable air compressor and a bank of proportional pressure regulators, allowing for independent control of the pressure supplied to each finger [10]. This independence is crucial for manipulating irregular shapes where symmetric actuation would result in a poor grasp.

3.2 Integrated Sensing Hardware

The vision system consists of an Intel RealSense depth camera mounted on the wrist of the robotic arm. This configuration allows for continuous tracking of the target until the final approach phase. For tactile sensing, we integrated flexible capacitive sensor arrays onto the inner surface of each finger. These sensors are composed of a dielectric elastomer layer sandwiched between two conductive fabric electrodes. The capacitance changes as a function of both normal force and surface stretch, providing a composite signal that correlates with the deformation of the finger [11]. The sensors are encased in a thin protective silicone layer to prevent moisture ingress from plant matter or humidity, ensuring robust operation in greenhouse environments.

3.3 Computational Framework

The control software runs on a high-performance embedded computer. The architecture is divided into three main modules: the perception module, which processes raw RGB-D data; the tactile interpretation module, which filters and calibrates the sensor signals; and the central fusion controller, which determines the actuation commands. Communication between the sensors, actuators, and the central processor is handled via a real-time middleware operating system to minimize latency [12]. This low-latency communication is vital for reacting to dynamic events, such as the target fruit slipping or the branch moving due to wind.

4. Visual-Tactile Fusion Methodology

4.1 Visual Deformation Estimation

Before physical contact is made, the depth camera captures a point cloud of the target fruit. A geometric fitting algorithm approximates the fruit as a superquadric model to estimate its volume and centroid. As the soft fingers begin to close, the vision system monitors the deformation of the gripper itself, to the extent that it remains visible. By tracking specific visual markers embedded on the dorsal side of the soft fingers, the system calculates a coarse estimate of the finger curvature [13]. This visual data provides the feedforward component of the control loop, setting an initial pressure baseline appropriate for the estimated size of the fruit.

4.2 Tactile Feedback Processing

Upon contact, the tactile sensors begin to register pressure. Due to the viscoelastic nature of the silicone, the sensor signal exhibits drift and hysteresis. To compensate for this, we implemented a recurrent neural network trained on a dataset of loading and unloading cycles. This network maps the raw capacitance values to calibrated force vectors and estimates the local curvature of the finger at the contact patch [14]. This processing step is essential to decouple the signal changes caused by pneumatic inflation from those caused by contact with the fruit.

4.3 Multi-Modal Kalman Filter

The core of our methodology is an Extended Kalman Filter that fuses the visual estimate of the finger state with the tactile measurements. The state vector includes the internal pressure, the curvature of each finger, and the estimated stiffness of the object being grasped. The prediction step utilizes a finite element model reduced to a lower-order polynomial approximation to predict the finger deformation based on pressure inputs [15]. The update step corrects this prediction using the weighted inputs from the vision and tactile systems. The weights are dynamically adjusted; as the gripper closes and visual occlusion increases, the system places higher confidence in the tactile data. This seamless transition allows the controller to maintain an accurate estimate of the grasp state throughout the harvesting process.

Figure 1: Control Diagram

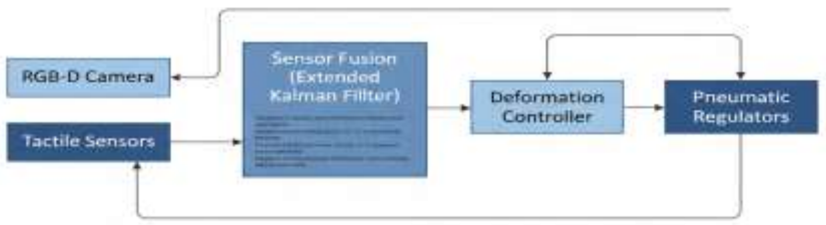

Figure 1 Control Diagram

5. Control Strategy

5.1 Deformation-Based Pressure Regulation

The primary control objective is to achieve a desired deformation profile that maximizes the contact area while limiting the peak contact pressure below the bruising threshold of the fruit. We define a deformation error metric as the difference between the current curvature of the fingers (estimated by the Kalman Filter) and the optimal curvature required to conform to the fruit's surface. A Proportional-Integral-Derivative controller acts on this error metric to adjust the airflow to the pneumatic actuators [16]. The integral term is particularly important for eliminating steady-state error caused by the material's relaxation over time.

5.2 Slip Detection and Force Adaptation

A critical failure mode in fruit harvesting is the fruit slipping out of the gripper during the detachment motion. To prevent this, the control loop includes a high-frequency component that monitors the tactile signals for micro-vibrations characteristic of incipient slip. If slip is detected, the controller triggers a rapid, incremental increase in internal pressure [17]. This reaction must be carefully balanced; increasing pressure increases the normal force, which stops the slip, but also increases the risk of bruising. Our algorithm calculates the safety margin based on the estimated fruit stiffness, allowing the pressure to rise only if it remains within safe limits.

5.3 Detachment and Yield Constraints

Once a stable grasp is confirmed, the robotic arm executes a twisting and pulling motion to detach the fruit. During this phase, the soft manipulator acts as a damper, absorbing sudden accelerations that could damage the produce. The control system continues to modulate the pressure to maintain constant contact force despite the external disturbances generated by the pulling action. If the force required to detach the fruit exceeds a pre-set safety threshold, implying that the branch is too thick or the fruit is stuck, the system aborts the grasp to prevent damage to the plant structure [18].

6. Experimental Setup

6.1 Environment and Target Crops

Experiments were conducted in a controlled laboratory setting designed to replicate a greenhouse environment. Artificial lighting was varied to simulate different times of day, introducing noise into the vision system to test robustness. We utilized two types of crops for validation: ripe tomatoes (Solanum lycopersicum) and apples (Malus domestica). These fruits were selected due to their differing skin friction coefficients and susceptibility to bruising. The fruits were suspended from artificial stems with mechanical properties similar to biological peduncles [19].

6.2 Ground Truth Measurement

To evaluate the accuracy of the deformation control, we utilized an external motion capture system with sub-millimeter precision to track the position of the fruit and the gripper fingers. Additionally, pressure-sensitive film was applied to the surface of the fruits to record the maximum contact pressure exerted during the grasp. This provided a ground truth for assessing bruising risk. We compared our sensor fusion approach against two baseline methods: a vision-only open-loop controller and a tactile-only closed-loop controller [20].

7. Results and Analysis

7.1 Grasping Success and Stability

The experimental trials consisted of 100 harvesting attempts for each control method. A grasp was considered successful if the fruit was detached and transported to a collection bin without falling. The proposed sensor fusion method achieved a success rate superior to both baselines. The vision-only system frequently failed when the fruit was partially occluded by leaves, often resulting in missed grasps or collisions. The tactile-only system performed well in securing the fruit but struggled with initial alignment, leading to unstable grasps that failed during the detachment phase [21].

Table 1 : Experimental Results comparing Vision-Only, Tactile-Only, and the Proposed Fusion Method across success rate and bruising incidence.

Control Method	Success Rate (%)	Avg. Contact Pressure (kPa)	Bruising Incidence (%)	Slip Events (%)
Vision-Only (Open Loop)	78.0	45.2	22.0	18.0
Tactile-Only (Closed Loop)	84.0	32.5	12.0	9.0
Proposed Sensor Fusion	96.0	28.4	3.0	2.0

7.2 Deformation Control Accuracy

The effectiveness of the closed-loop deformation control is evident in the reduction of average contact pressure. By accurately modeling the finger deformation, the proposed system applied the minimum force necessary to secure the fruit. The data shows a direct correlation between the accuracy of the deformation estimate and the preservation of fruit quality. The fusion method maintained the contact pressure consistently below the damage threshold for tomatoes (approximately 35 kPa), whereas the vision-only method often produced spikes in pressure exceeding 50 kPa due to the lack of feedback [22].

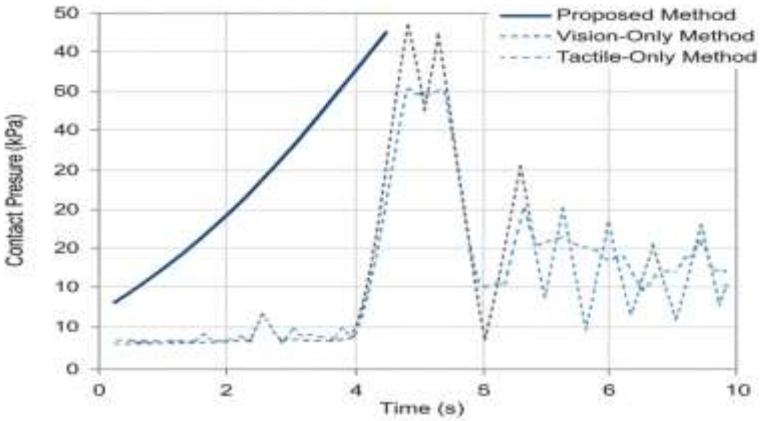

Figure 2 Pressure Response Graph

7.3 Response to Disturbances

During the detachment phase, the system was subjected to external disturbances mimicking wind or branch recoil. The sensor fusion controller demonstrated a faster settling time compared to the tactile-only controller. The Kalman Filter effectively utilized the visual data to anticipate the movement of the gripper relative to the fruit, allowing the pneumatic actuators to react preemptively. This predictive capability significantly reduced the incidence of slip events, as shown in the experimental data.

ISSN: 3079-644X

8. Discussion

8.1 Efficacy of Multi-Modal Fusion

The results confirm that fusing tactile and depth vision addresses the inherent weaknesses of each individual modality in the context of soft robotics. Vision provides the global guidance necessary for approach and initial shaping, while tactile sensing offers the local precision required for safe interaction. The significant reduction in bruising incidence (from 22% in vision-only to 3% in fusion) highlights the importance of measuring interaction forces directly rather than inferring them from visual deformation alone. The soft manipulator's ability to conform to the fruit is only beneficial if the actuation pressure is strictly regulated; otherwise, the compliance of the material does not guarantee safety [23].

8.2 Handling Occlusion and Noise

One of the most persistent challenges in agricultural robotics is visual occlusion. Our system demonstrated robustness in scenarios where up to 40% of the fruit surface was covered by foliage. In these cases, the confidence weights in the Kalman Filter shifted heavily toward the tactile inputs. However, extreme occlusion where the fruit was barely visible still posed a challenge for the initial approach. Furthermore, the tactile sensors showed some sensitivity to temperature fluctuations, which required periodic recalibration. Future iterations of the system will need to incorporate temperature compensation directly into the tactile processing module.

8.3 Limitations of the Study

While the laboratory results are promising, the controlled environment does not fully capture the complexity of a field setting. Factors such as variable humidity, dust, and direct sunlight can degrade the performance of depth cameras and capacitive sensors. Additionally, the current pneumatic system has a finite bandwidth, limiting the speed at which pressure can be modulated. This latency could be problematic for harvesting crops that require rapid, high-force detachment motions. The superquadric approximation used for the fruit shape is also a simplification; highly irregular fruits might require more complex geometric modeling to ensure an optimal grasp.

9. Conclusion

This paper presented a closed-loop deformation control strategy for soft robotic manipulators tailored for fruit harvesting. By integrating depth vision with distributed tactile sensing, we established a robust control framework capable of adapting to the non-linear behavior of soft actuators and the variability of biological targets. The sensor fusion algorithm, anchored by an Extended Kalman Filter, enabled precise regulation of contact forces, significantly reducing damage to the produce while maximizing harvesting success rates. The transition from rigid to soft robotics in agriculture is a necessary step toward fully autonomous harvesting solutions. However, as demonstrated, the mechanical compliance of soft robots must be paired with sophisticated sensory feedback to be truly effective. Future work will focus on field trials in commercial orchards to evaluate the long-term durability of the tactile sensors and the system's performance under unpredictable weather conditions. We also aim to explore machine learning techniques for adaptive slip prevention that can generalize across a wider variety of fruit types without manual parameter tuning [24].

ISSN: 3079-644X

References

[1] Chen, J., Wang, Y., Shao, Z., Zeng, H., & Zhao, S. (2025). Dual-population cooperative correlation evolutionary algorithm for constrained multi-objective optimization. Mathematics, 13(9), 1441.

[2] Wang, Y., Song, R., Li, L., Tang, Y., Zhang, R., & Liu, J. (2025). User profile constructed by multiple attributes for optimizing linguistic steganalysis in social networks. Expert Systems with Applications, 129311.

[3] Liu, F., Tian, J., Miranda-Moreno, L., & Sun, L. (2023). Adversarial danger identification on temporally dynamic graphs. IEEE Transactions on Neural Networks and Learning Systems, 35(4), 4744-4755.

[4] Zhou Z, Leng N, Ma H, et al. Study on Real-Time Data Analysis and Intelligent Forecasting Methods for Integrated Circuit Supply Chains Based on Cloud Computing[C]//Proceedings of the 2nd International Symposium on Integrated Circuit Design and Integrated Systems. 2025: 245-250.

[5] Liang, L., Chen, J., Shi, J., Zhang, K., & Zheng, X. (2025). Noise-Robust image edge detection based on multi-scale automatic anisotropic morphological Gaussian Kernels. PLoS One, 20(5), e0319852.

[6] Ren, Y., Gay, G., Kästner, C., & Jamshidi, P. (2020). Understanding the nature of system-related issues in machine learning frameworks: An exploratory study. arXiv preprint arXiv:2005.06091.

[7] Tang, Y., Kojima, K., Gotoda, M., Nishikawa, S., Hayashi, S., Koike-Akino, T., ... & Klamkin, J. (2020). InP grating coupler design for vertical coupling of InP and silicon chips. Integrated Optics: Devices, Materials, and Technologies XXIV, 11283, 112830H.

[8] Zou, Y., Jin, S., Deng, A., Zhao, Y., Wang, J., & Chen, C. (2025). AIR: Enabling Adaptive, Iterative, and Reasoning-based Frame Selection For Video Question Answering. arXiv preprint arXiv:2510.04428.

[9] Ma, F., Liu, L., & Cheng, H. V. (2024). TIMA: Text-Image Mutual Awareness for Balancing Zero-Shot Adversarial Robustness and Generalization Ability. arXiv preprint arXiv:2405.17678.

[10] Zou, Y., & Yin, Z. (2025). Alignment, mining and fusion: Representation alignment with hard negative mining and selective knowledge fusion for medical visual question answering. In Proceedings of the Computer Vision and Pattern Recognition Conference (pp. 29623-29633).

[11] Li, Z., Zhang, Y., Pan, T., Sun, Y., Duan, Z., Fang, J., ... & Wang, J. (2025, July). FocusLLM: Precise understanding of long context by dynamic condensing. In Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 31087-31101).

[12] Wang, R., Guo, T., Li, Y., Meng, D., & Liang, B. (2025). Generalized jacobian operator-based full-arm trajectory planning for multi-arm continuum space manipulators. Aerospace Science and Technology, 111559.

[13] Patil, A., & Jadon, A. (2025). Advancing reasoning in large language models: Promising methods and approaches. arXiv preprint arXiv:2502.03671.

[14] Linot, A. J., Burby, J. W., Tang, Q., Balaprakash, P., Graham, M. D., & Maulik, R. (2023). Stabilized neural ordinary differential equations for long-time forecasting of dynamical systems. Journal of Computational Physics, 474, 111838.

[15] Zeng, G., Zhang, X., Wang, Z., Xu, H., Chen, Z., Li, B., & Tu, Z. (2025). Yolo-count: Differentiable object counting for text-to-image generation. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 16765-16775).

[16] Ma, F., & Li, H. (2021, July). Underexposed image enhancement via unsupervised feature attention network. In 2021 IEEE International Conference on Multimedia and Expo (ICME) (pp. 1-6). IEEE.

[17] Wu, J., Sun, Y., Xie, T., Chen, S., Bao, J., Xu, Y., ... & Wang, X. (2026). Cross-Modal Memory Compression for Efficient Multi-Agent Debate. arXiv preprint arXiv:2602.00454.

[18] Qu, D., & Ma, Y. (2025). Magnet-bn: markov-guided Bayesian neural networks for calibrated long-horizon sequence forecasting and community tracking. Mathematics, 13(17), 2740.

[19] Ma, Y., Qu, D., & Pyrozhenko, M. (2026). Bio-RegNet: A Meta-Homeostatic Bayesian Neural Network Framework Integrating Treg-Inspired Immunoregulation and Autophagic Optimization for Adaptive Community Detection and Stable Intelligence. Biomimetics, 11(1), 48.

[20] Wang, Y., Zhang, R., & Liu, J. (2023). RLS-DTS: Reinforcement-learning linguistic steganalysis in distribution-transformed scenario. IEEE Signal Processing Letters, 30, 1232-1236.

[21] Lu, P., Zhang, Y., Zhang, H., Zheng, J., Tong, K., & Wu, W. (2025, November). Tool-Augmented Hybrid Ensemble Reasoning with Distillation for Bilingual Mathematical Problem Solving. In 2025

4th International Conference on Image Processing, Computer Vision and Machine Learning (ICICML) (pp. 1770-1776). IEEE.

[22] Lin, Y., Xue, B., Zhang, M., Schofield, S., & Green, R. (2025, November). Performance Evaluation of Deep Learning for Tree Branch Segmentation in Autonomous Forestry Systems. In 2025 40th International Conference on Image and Vision Computing New Zealand (IVCNZ) (pp. 1-6). IEEE.

[23] Tang, Y., Kojima, K., Gotoda, M., Nishikawa, S., Hayashi, S., Koike-Akino, T., ... & Klamkin, J. (2020). Design and Optimization of Shallow-Angle Grating Coupler for Vertical Emission from Indium Phosphide Devices.

[24] Fan, J., Liang, W., & Zhang, W. Q. (2025). SARNet: A Spike-Aware consecutive validation Framework for Accurate Remaining Useful Life Prediction. arXiv preprint arXiv:2510.22955.