FruitTouch: A Perceptive Gripper for Gentle and Scalable Fruit Harvesting

Ruohan Zhang[1], Mohammad Amin Mirzaee[1], and Wenzhen Yuan[1]

***Abstract*— The automation of fruit harvesting has gained increasing significance in response to rising labor shortages. A sensorized gripper is a key component of this process, which must be compact enough for confined spaces, able to stably grasp diverse fruits, and provide reliable feedback on fruit condition for efficient harvesting. To address this need, we propose FruitTouch, a compact gripper that integrates high-resolution, vision-based tactile sensing through an optimized optical design. This configuration accommodates a wide range of fruit sizes while maintaining low cost and mechanical simplicity. Tactile images captured by an embedded camera provide rich information for real-time force estimation, slip detection, and softness prediction. We validate the gripper in real-world cherry tomato harvesting experiments, demonstrating robust grasp stability, effective damage prevention, and adaptability to challenging agricultural conditions.**

I. Introduction

The agricultural robotics system for harvesting has been a focus for the community due to the lack of human labor. When designing the automated harvesting system, the choice of the end effector is critical since it is directly responsible for handling delicate fruits, preventing slips, inferring fruit firmness, and preserving the quality of the fruits by avoiding crushing or rubbing during the harvest process [1]. Existing agricultural end-effectors commonly rely on visual feedback for fruit detection [2] and mechanical designs such as suction cups [3], [4], scissor cutters [5], or multi-finger grippers [6], [7] to enhance harvesting success rate. While effective under ideal conditions, these solutions face several limitations. Vision-based feedback can be unreliable when the fruit is occluded by foliage or when lighting conditions vary. Purely mechanical end-effectors, such as suction cups or cutting-based tools (e.g., scissors or blade mechanisms), often lack robustness to environmental variability, including fruit size variation, clustered growth, or surface wetness. While some of these systems later augment vision with low-resolution or binary fingertip contact sensors to signal contact onset or threshold events [8], [9], such signals provide little spatial or directional information. As a result, they cannot recover the rich contact state and fruit material properties for achieving reliable grasp.

On the contrary, human pickers excel at this task by integrating tactile and force cues to detect contact state and assess ripeness during manipulation. Recent advances in high-resolution tactile sensing, such as GelSight [10]–[13] technology, offer a promising path toward overcoming

[1]Ruohan Zhang, Mohammad Amin Mirzaee, and Wenzhen Yuan are with University of Illinois at Urbana-Champaign, Champaign, IL, USA {rz21,mirzaee2,yuanwz}@illinois.edu

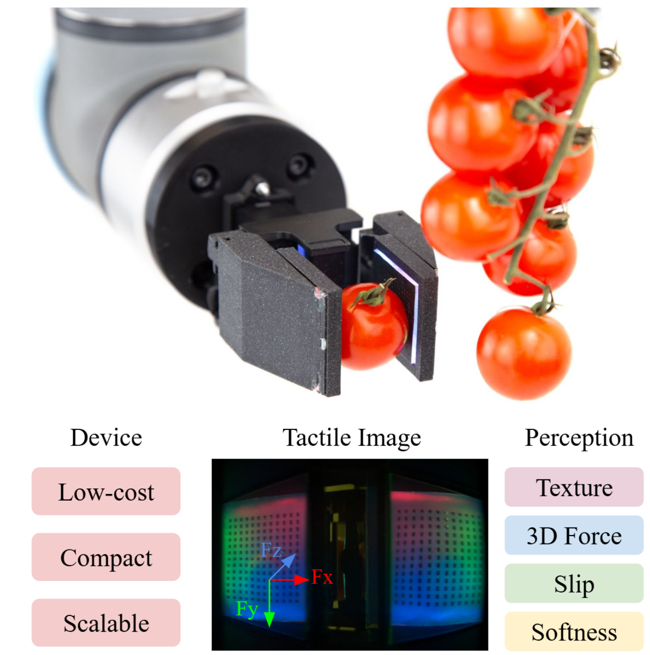

Fig. 1. Demonstration of the FruitTouch gripper harvesting cherry tomatoes. The hardware design is optimized for compactness, low cost, and scalability, while the perception system measures high-resolution contact geometry, 3D force, slip, and object softness. Together, the design integrates mechanical efficiency with rich tactile sensing to enable reliable and efficient fruit harvesting.

these limitations. GelSight sensors capture detailed surface geometry and deformation patterns, enabling accurate estimation of contact forces [14], [15], detection of slip [16], [17], and assessment of surface compliance [18]. These capabilities have driven successful applications of GelSight in various robotic manipulation tasks; however, its integration into agricultural grippers remains limited.

In this work, we present FruitTouch, a compact and lightweight agricultural end-effector that integrates high-resolution tactile sensing to enable robust fruit perception and harvesting in real-world scenarios. The design adopts a parallel-jaw mechanism with a single camera to sensorize both gripping surfaces, reducing cost and complexity for fabrication. Its thin, wedge-shaped fingers allow the gripper to reach into dense foliage and clutter, which is critical for practical field deployment. Besides, the scalable mechanical and optical design opens up the possibility of accommodating fruit sizes ranging from fruits even smaller than cherry tomatoes (∼28 mm diameter) to large fruits like apples (∼75 mm diameter). To summarize, our proposed FruitTouch gripper can:

- Reconstruct fruit surface texture with high precision,
- Detect and prevent fruit crushing through accurate 3D force measurement,
- Identify and respond to slip events during grasping, and
- Classify fruit softness as a proxy for ripeness.

We use the example of cherry tomatoes as the target fruit —a soft, vine-grown fruit typically forming clusters of around 12 fruits per bunch, with an average fruit diameter of approximately 28.3 mm [7]. This presents significant challenges for common end-effectors due to the need for gentle yet secure handling in cluttered and constrained environments. By integrating high-resolution tactile sensing, a compact and scalable mechanical and optical design, and adaptability for agricultural settings, our system demonstrates strong performance in these demanding conditions, achieving high precision and operational efficiency during harvesting. We believe this work can contribute toward enabling scalable, automated harvesting solutions that reduce harvest losses, increase harvesting efficiency, and lessen reliance on human labor in the foreseeable future.

II. Related Work

A. Agricultural End-Effectors for Harvesting

According to the detachment method, agricultural end-effectors can be broadly classified into two categories: cutting-based and grasping-based [19]. Cutting-based methods use a blade or knife to sever the fruit from the stem, making them suitable for fruits with relatively long and accessible stems. For example, [20], [21] employed self-designed cutting end-effectors combined with suction cups to harvest strawberries and tomatoes, while [22] proposed a cut-clip mechanism for detaching grape clusters.

In contrast, grasping-based methods harvest by directly contacting the fruit's surface with the end-effector and applying sufficient wrench for detachment. For example, [23] presented a multi-arm kiwifruit harvesting robot equipped with multiple clamping grippers, achieving an 84% harvesting success rate with an average cycle time of 5.5 s per fruit in field trials. [24] developed a strawberry picking system using a soft pneumatic gripper, which achieved a 78% success rate and 23% damage rate. These grasping-based approaches not only resemble the way humans hold fruit with their fingers but also offer the opportunity to directly sense the fruit through surface contact. This tactile interaction when grasping can provide valuable information about the fruit's physical properties, enabling more informed control strategies to improve harvesting success while reducing damage. However, effectively exploiting this interaction requires suitable sensing technology, for which vision-based tactile sensors offer a promising solution.

B. Perception Using Vision-based Tactile Sensors

Vision-based tactile sensors, particularly GelSight [25], have shown great potential for capturing detailed contact geometry and providing rich feedback for robotic manipulation tasks. Such sensors can enable accurate force estimation, slip detection, and object softness detection, which make it suitable for intergration into agriculture end-effectors.

For example, [10] demonstrated that the relationship between indentation volume and normal force is approximately linear. Building on this insight, subsequent works [26], [27] employed finite element methods (FEM) for more refined force estimation. Shear force has been estimated by tracking the displacement of surface markers [10], [15], which have also been leveraged for slip detection: average marker displacement has been used for small objects [28], while entropy-based measures have been applied to larger contact areas [16]. Object softness prediction has likewise been an active area of research, with early approaches modeling softness as a linear function of image brightness [29], and more recent studies incorporating analytical models to better interpret tactile signals and improve compliance estimation [30].

Despite their potential for enhancing perception, existing vision-based tactile systems often suffer from bulky form factors, high cost, limited capability, or poor generalizability, making them difficult to apply directly to harvesting. For example, [31] presented a GelSight-equipped 5-DoF gripper for in-hand manipulation, but its complex multi-DOF structure runs counter to the simplicity and scalability required in agricultural applications. Likewise, [18] used GelSight as a standalone post-harvest firmness tester rather than integrating it into the picking process, while [32] developed a strawberry-harvesting end-effector that, although capable of force feedback, remained bulky and lacked both slip detection and ripeness estimation.

In contrast, our proposed gripper is sensorized, compact, and low-cost, providing integrated perception of force, slip, and ripeness within a single harvest-ready design. Unlike prior GelSight-based sensors for agriculture, it employs a single camera shared across both fingers through a common optical path. This approach reduces cost while maintaining sensing effectiveness under varying gripper configurations, enabling practical deployment in dynamic harvesting environments.

III. Sensor-Integrated Gripper Design

To achieve a practical balance between simplicity, robustness, and sensing capability, we adopt a parallel-jaw gripper design, which reduces mechanical complexity, minimizes interference with both the optical system and the surrounding environment during harvesting. This also mimics the natural two-finger picking strategy, with integrated tactile sensing on both fingers. The following subsections present the mechanical specifications of the gripper in Section III-A and the optical design in Section III-B.

A. Mechanical Design

We placed the racks and gear at the base of the gripper to actuate the fingers, to minimize interference with the optical subsystem. Racks are attached to the fingers and coupled with a gear connected to a DYNAMIXEL XC330-M288-T motor, as shown in Figure 2A.