Computers and Electronics in Agriculture 213 (2023) 108202

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

A soft, sensorized gripper for delicate harvesting of small fruits

Francesco Visentin *, Fabio Castellini, Riccardo Muradore

University of Verona, Department of Engineering for Innovation Medicine, Strada le Grazie, 15, Verona, 37134, Verona, Italy

ARTICLE INFO

Keywords:
Soft grippers
Precise picking
Autonomous harvesting
Small fruit harvesting

ABSTRACT

Harvesting fruits and vegetables is a complex task worth to be fully automated with robotic systems. It involves several precision tasks that have to be performed with accuracy and the appropriate amount of force. Classical mechanical grippers, due to the complex control and stiffness, cannot always be used to harvest fruits and vegetables. Instead, the use of soft materials could provide a visible advancement. In this work, we propose a soft, sensorized gripper for harvesting applications. The sensing is performed by tracking a set of markers integrated into the soft part of the gripper. Different machine learning-based approaches have been used to map the markers' position and dimensions into forces in order to perform a close-loop control of the gripper. Results show that force can be measured with an error of 2.6% in a range from 0 to 4 N. The gripper was integrated into a robotic arm having an external vision system used to detect plants and fruits (strawberries in our case scenario). As a proof of concept, we evaluated the performance of the robotic system in a laboratory scenario. Plant and fruit identification reached a positive rate of 98.2% and 92.4%, respectively, while the correct picking of the fruits, by removing it from the stalk without a direct cut, achieved an 82% of successful rate.

1. Introduction

Nowadays, automation plays a major role in the farming industry, helping in minimizing the waste of products, the time needed to accomplish complete tasks, and optimizing the crop production cycle (Lowenberg-DeBoer et al., 2020). Among the main tasks in agricultural processes, the ones involving the manipulation of fruits and vegetables continue to be among the most time-consuming and labor-intensive, resulting in low efficiency and limited competitiveness (Navas et al., 2021). The situation was also amplified after the COVID-19 crisis, as the pandemic increased the labor shortages of seasonal workers unable to travel between regions and be housed during the harvesting seasons. For these reasons, researchers, and companies focused their efforts on finding solutions to automate these manual operations combining multidisciplinary fields such as biological science, control engineering, robotics, and artificial intelligence.

Harvesting fruits and vegetables is a complex task to be fully automated with a robotic system. It involves precision in fruit identification, real-time decision-making about harvesting and storing, and delicate handling of both the plant and the harvested product. In manual harvesting, humans use their hands to move different elements of plants, and grasp and detach the fruits, either directly or with the help of a tool. Attempts to emulate human skills during harvesting have resulted in numerous mechanical end-effectors that vary the number of integrated fingers (Zhang et al., 2020). Having multiple fingers, as proposed in Ozawa and Tahara (2017), King et al. (2018), Mizushima et al. (2018), Vulliez et al. (2018), provides high dexterity and compliance during the task, with the downside of being fragile and difficult to control due to the large number of degrees of freedom (DoFs). On the contrary, the simple mechanical structure of the parallel grippers—which just have two fingers—simplifies the control but reduces the adaptability during grasping. An alternative solution to rigid, anthropomorphic grippers can be found in the use of deformable materials exploited in the field of soft robotics. Soft robotics research focuses on the development of novel machines able to safely interact with the environment by providing novel design, control strategies, and manufacturing techniques to develop robots made of compliant materials (Ahmed et al., 2022).

Soft robots, not only aim at being safer for humans but may also solve different open challenges in robotics. Being compliant, they are flexible, harder to break or damage, and adaptable to unstructured and dynamic environments. By replacing the intricate rigid body joint mechanics with simple compliant mechanisms, the number of parts required is significantly reduced, leading to lower costs for maintenance and assembly, and a simpler control architecture. They do not require an in-depth characterization of the object to handle, ensuring good performance even with minimal, if not any, force feedback—which is not possible by using traditional robotic approaches. In addition,

* Corresponding author.
E-mail address: francesco.visentin@univr.it (F. Visentin).

https://doi.org/10.1016/j.compag.2023.108202
Received 1 April 2023; Received in revised form 19 July 2023; Accepted 30 August 2023
Available online 12 September 2023

they can meet hygiene and strict manipulation requirements while operating with delicate or fragile products. These properties make them a promising solution for the development of grippers for handling fruits and vegetables (Hughes et al., 2016).

The field of soft robotics already counts several solutions for food handling and manipulation (Elfferich et al., 2022), however, most of those are not specifically designed for picking fruits and vegetables (Navas et al., 2021). Among all the solutions, the ones based on fluidic elastomer actuators are the most promising for better adapting to cope with the different sizes of medium to large agricultural products. The classical approaches to soft, pneumatic grippers are all based on the PneuNet design, which consists of a series of channels and chambers inside an elastomer, which expand toward the least stiff region when pressurized (Ilievski et al., 2021).

In an early work, Wang and Hirai (2018) presented a soft gripper optimizing such design to perform precise food handling. Visentin et al. (2021) extended the concept also by providing selective stiffening to improve the grasping capabilities of the gripper. By increasing the number of chambers, and by arranging four fingers in a circular configuration, Kuriyama et al. (2019) developed a gripper able to grasp small fruits. By changing the design from longitudinal to radial, Wang et al. (2021) obtained a hemispherical griper that can be selectively inflated to provide better adhesion to the object. It has been proved that the gripper can withstand payloads up to 0.5 kg. A similar solution was also proposed by Galley et al. (2019). An alternative solution to the use of anthropomorphic design is the use of suction and the use of vacuum pumps. Wang et al. (2020) proposed a dual-mode soft gripper combining grasping and suction abilities located at each finger: the design allows handling objects of different shapes and weights by using either of the grasping mechanisms.

Different from direct casting, the use of 3D printers to develop soft grippers speeds up the development process and ensures a high level of repetitiveness in the manufacturing. Blanes et al. (2014), for example, proposed a set of fully 3D-printed, fruit-specific, monolithic designs for soft grippers. Liu et al. (2018), instead, only focused on the development of an underactuated, soft robotic gripper with compliant fingers for grasping fruits. Tawk et al. (2019), then extended the concept to the development of an omni-purpose soft gripper. The advantage of their approaches is the fast manufacturing that is not always ensured in the previously presented solutions. Recent solutions, such as the one presented by Anon (2023), not only reduce the amount of material used but are designed to enable passive adaptation to the different sizes of the harvested product.

All the presented solutions do not integrate a sensing mechanism to capture the applied forces. As previously stated, the use of compliant materials helps in reducing the need for precise force control when handling delicate products like berries. Solutions for the specific task of harvesting strawberries (Arima et al., 2004; Feng et al., 2012; Xiong et al., 2018; Anon, 2019, 2023) are mostly based on the use of a knife to cuts off the fruit at the stalk level, then the fruit is collected and stored. This approach does not require holding the fruit still, thus force feedback is not needed. However, it can increase the possibility to damage the fruit during further processing. To avoid such an issue, a different type of approach should be used to grasp and detach the fruit from the plant.

Starting from this problem, in this work, we propose a soft, sensorized gripper that can be mounted on a robotic manipulator to perform strawberry picking. We tested its performance in a complete workflow: from the single plant identification, the selection of the mature fruit, up to its harvesting. Differently from the previously presented solutions, our approach does not remove the fruit by cutting its stalk, instead, it performs a clean removal as human do. The paper is structured as follows: in Section 2 we will introduce the manufacturing and the working principle of the gripper, the method used to estimate the interaction force, and the external vision system used for the plant/fruit detection. In Section 3, we will first present the characterization of the sensorized gripper, then the evaluation of its use on a robotic system in a laboratory scenario. Then, we will conclude with Section 4 highlighting the advancement of the proposed system and future improvements.

2. Materials and methods

In this work, we focused on the development of a cheap, 3D-printed, soft gripper capable of sensing forces when in contact with the external environment. This was done by manufacturing a patterned, hemispherical, soft dome and tracking the displacement of every single marker in the pattern under different conditions. Similar solutions have already been proposed, however, most of the research on this topic has been focusing on soft but flat tactile sensors. An example is the work of Zhang et al. (2019) where the contact force and the torque have been estimated by a commonly used motion analysis technique which is based on vector field analysis estimating the position of divergence source, sink, and vertex of rotational motion. The data acquired using a highly dense pattern are fed into a multi-layer perceptron (three layers) which provides a correct estimation of the forces with a total error of 2 N (on average) for the normal and torsional forces, and more than 5 N for the torsional one. Other planar sensors and their use for slip detection are presented and discussed in Chen et al. (2018).

The most promising solutions for soft sensing, such as the ones proposed by Alspach et al. in Alspach et al. (2019) and then by Kuppuswamy et al. (2020), are based on thin, patterned membranes and the use of depth cameras. The sensors can generate a highly compliant dense geometry of the membrane that can be used to perform object, shape, and texture classification (using a Deep Neural Network), object sorting, object pose estimation, and tracking. Even if promising, those cameras require a minimum sensing distance of 100 mm, which restricts the scalability of the proposed solution to more compact designs. Classical RGB cameras have also been used to track the motion of markers. An early work on the topic is presented by Sakuma et al. in Sakuma et al. (2018), where a universal gripper (Brown et al., 2010) was modified to integrate sensing by measuring the deformation of the external membrane using optical sensing. The prototype could identify and grasp cylindrical and rectangular objects with lengths in the range from 10 to 70 mm. Later, Scharff et al. (2022) improved the technique by increasing the number and color of the patterns. The sensing principle is based on the relative displacement between differently colored markers that lie on two separate layers. This approach helped to encode the distance between markers using a monocular camera, forcing the normal deformation to be estimated from the lateral displacement of the markers. It has a higher sensing resolution than previous solutions, however, it was only used to perform accurate curvature estimations when the sensor is pressing against a positively or negatively curved object. Other solutions based on hemispherical domes abandoned the markers to exploit optical features induced by reflecting coatings applied to the domes. In Fernandez et al. (2021), Fernandez et al. used a rigid, acrylic dome as a waveguide covered by a silicone cap. Eight LEDs act as fiducial markers and the light injected into the acrylic dome—which acts as a waveguide—is internally reflected except where the cap contacts it. This is exploited to easily sense either single or multi-contact with the external environment. Do and Kennedy (2022), instead, proposed a compact, high-resolution tactile sensor capable of reconstructing the shape of the object in contact. This was possible by the use of a highly-reflective coating on the internal part of the dome, and a calibration process (using 3D printed objects and inferring CAD models) based on photometric stereo algorithms and Deep Neural Network. Both solutions are promising, however, they require special chemical treatments that might not be compatible with the crops.

While the previous works have been focusing on the characterization of the sensors only, Sakuma et al. (2019) also integrate the sensor in a parallel gripper. The tactile unit was developed using direct casting of the external membrane which was then manually polished

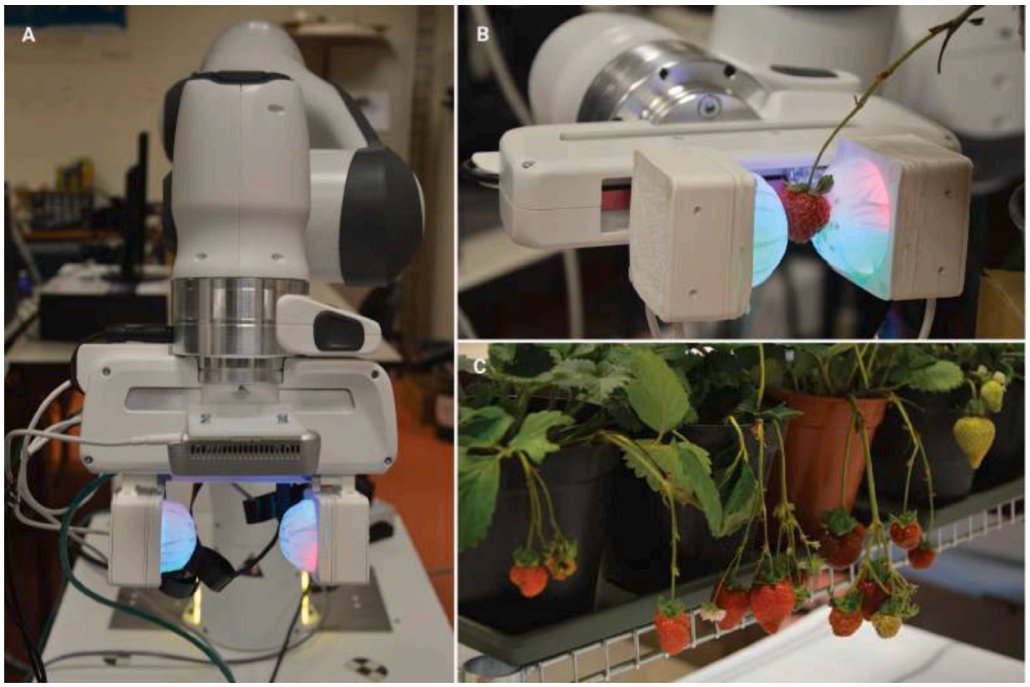

Fig. 1. (A) The proposed soft gripper is integrated into the robotic system. The gripper of the Franka Emika Panda robot has been modified to host two soft, sensorized fingers developed to gently grasp delicate objects. Two additional structures have been added to the gripper: a support for the RGB-D camera (Intel Realsense D345i, Intel, USA), and an additional holder for the electronics (Raspberry Pi 3 Model B) used to acquire the data from the sensorized gripper. (B) The real-time force estimation during a grasping task. (C) A picture of the experimental setup used in this paper to validate out approach.

to increase its transparency. The markers were manually painted over the internal side of the membrane which was also filled with acrylic beads and a mixture of silicon and paraffin oil. No light was used to increase the contrast of the markers; thus, the system can only be used in a controlled environment with no change in the light condition (e.g., open field). The main advantage is the integration of a jamming transition that can be used to maintain a stable grasp. However, it is difficult to reset the gripper to its natural state after releasing the object, making it more complicated to plan and execute successive picking (Amend et al., 2016). A more advanced configuration was proposed by Choi et al. in Choi and Tahara (2020) where a similar vision-based tactile unit was integrated in a multi-fingered robotic hand. The tactile unit was directly cast using a silicon rubber and the marker was manually painted. The proposed sensor has a compact footprint (radius, 30 mm) and can measure normal forces in the range of 4 to 6 N which are not suitable for handling delicate fruits such as berries.

Different from the previous research, our approach aims to reduce the complexity of the manufacturing process while preserving the capabilities to accurately measure forces in real-time during a task of grasping. Previous research, on the contrary, mostly focused on the shape reconstruction of the object in contact, or the detection of shear forces as feedback for more precise control of the gripper. In addition, the full integration with a robotic system provides additional improvements from the current state of the art. Fig. 1 shows the developed gripper mounted on the robotic system, the test scenario, and the result of a grasping task.

2.1. Manufacturing of the sensing units

Prioritizing ease of manufacturing and reproducibility, we decided to avoid the use of multi-material additive manufacturing and molding and focus on 3D printing. The main body of the gripper is the one provided by the Franka Emilka Panda robot, from which we substituted the original fingers with the developed sensing devices.

Each of the sensing devices is based on a 3D-printed, soft, hemispherical dome, and a 5MP, RGB camera (OV5647, Tangxi, China) with a 175° fish-eye lens. To ensure the proper illumination, a ring of twelve RGB LEDs (NeoPixel Ring 12 RGB, Adafruit, USA) was placed beneath the dome. A case, made of PLA (Polylactic Acid), was printed using an FDM (Fuse Deposition Modeling) printer (X1-Carbon, Bambulab, China) and used as a protective shield for the main components. It has been designed considering a modular approach to decouple as much as possible the different layers and simplify the assembly. A second holder is placed on the backside of the gripper to host the electronics (Raspberry Pi 3 Model B) needed to acquire the data from the cameras and to control the illumination of the LEDs. Fig. 2 shows the gripper and its components.

Each of the soft, hemispherical domes was 3D printed using a stereo lithography 3D printer (Form 2, Formlabs, USA) and a proprietary polymeric resin (Elastic 50 A, Formlabs, USA). The material supports a maximum printing resolution of 100 μm and it is designed to achieve a 50 A shore hardness when cured. After the printing, the domes were first cleaned in IPA (Isopropyl Alcohol) for 10 minutes, then cured in a dedicated ultraviolet oven for about 20 min at 60°. During the process, when half-cured, the patterns on the internal part of the dome were filled with the same uncured material (Elastic 50 A, Formlabs, USA), to which was added a black pigment (Black color kit, Formlabs, USA) to maximize the contrast. Then, the domes were again placed in the oven to terminate the curing process.

After the second curing, the domes can be integrated into the gripper. To avoid alteration of the images captured by the cameras as a consequence of stains left by strawberry juice or dust, we covered the domes with a thin layer (0.03 mm) of white latex which has the following effects on the sensing system: (i) it improves the contrast of the dots by reducing the light from the external environment (the dome is semi-transparent, and light can easily pass through it); (ii) it reduces the effects of cast shadows when the gripper gets in contact with an object (i.e., the contact point is a blend of the colors, not a black area); and (iii) it provides an easily, food-grade, replaceable interface that can be removed in case of need.

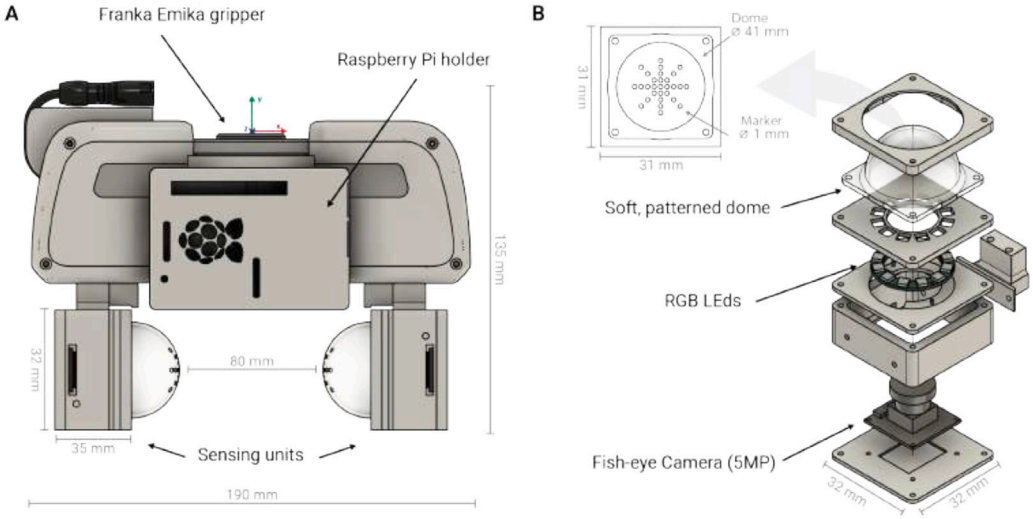

Fig. 2. The sensorized gripper and its components with relative dimensions. (A) A schematic representation of the gripper with the two sensing units connected and the electronic holder. (B) The exploded view of each unit with the featured main components. On top, a highlight on the soft dome shows the pattern used in this work.

The final dimensions of each sensing unit are $32 \times 32 \times 35$ mm^3, not considering the dome. The dome has a square base of 31×31 mm^2, a diameter of 41 mm, and a wall thickness of 1.5 mm. When mounted to the gripper (weight 0.7 kg), the total mass o the assembly becomes 1.12 kg, which is within the maximum payload of the robotic arm (3 kg) leaving room for additional weight.

2.2. Working principle of the soft gripper

The main idea behind the approach is the capability to identify and track the location of a pre-defined pattern and map it into the estimation of the interaction force. As a pattern, we used a double cross design with 29 individual markers. Each dot has a diameter of 1 mm, and it is engraved for 0.75 mm inside the hemispherical dome. We made this decision to maximize the detection rate and tracking reliability of each marker. The pattern was designed in such a way that diagonal crosses can be distinguished from orthogonal crosses by the number of markers (9 on the vertical arms, 7 on the diagonals) and the different spacing between the markers (1 mm and 2.4 mm, respectively). To further distinguish the direction, we set the LEDs to different pure colors (i.e., red, green, and blue) which can be easily segmented with classical computer vision approaches.

Each camera is connected to a Raspberry Pi 3 Model B which acquires and then scales down the images before sending them to the main computer. Then, the images are converted to grayscale and thresholded to maximize the contrast of the dark markers with respect to the rest of the image. In each binary image, the connected white pixels are grouped forming blobs, which are then merged if too close to each other. To avoid erroneous identification of the markers, we removed all the blobs having the centroid outside a radius equal to two-thirds of the frame length. The binary image is then processed to extract its contours, which are then fitted using the Fitzgibbon et al. (1999) algorithm. The output of this process is a set of coordinates and radii relative to the different markers. Fig. 3 shows the results of the marker identification. The proposed approach is robust (detection rate 98.7%) even when high deformations are involved (≥4 N). In this specific case, the shape of the marker in the images changes from circular to elliptic. To correctly measure the radius of the marker, the diameter of the circular marker can be computed as twice the semi-major axis of the ellipse shown in the camera image. The semi-major and semi-minor axes can then be computed using the ellipse fitting algorithm proposed in Fitzgibbon et al. (1999). However, for smaller deformations, the shape transition from a circle to an ellipse is less pronounced.

Once the markers are properly detected, their positions are tracked over time. For every received frame (frame rate, 20 fps), marker detection is applied, and the new coordinates are associated with the closer marker of the previous frame. Using the Euclidean distance as a metric to sort markers, we obtained robust and replicable results. To further improve the reliability, we created a Boolean interpolation matrix that contains the tracking state (i.e., true or false) of the 29 markers in each frame. Once all the states are collected, it is possible to retrieve the positions and radii of the lost markers using linear interpolation. This was particularly useful in the off-line training of the Neural Networks used to estimate the forces. Fig. 4 shows the effects of a large deformation (4 N) applied in the center of the dome.

2.3. AI-based force estimation

We exploited several approaches to map the displacement of the markers into force data (acquired with an external load cell). As in Yamaguchi and Atkeson (2017), Yamaguchi (2018), we first approached the problem with a linear approximation based on the formula of the elastic force. For each marker, we computed the total displacement, (d_x, d_y), and we estimated the force along the three axes as a linear combination of that values as follows:

$$[f_x, f_y, f_z] = \left[c_x d_x, c_y d_y, c_z \sqrt{d_x^2 + d_y^2}\right], \tag{1}$$

where (c_x, c_y, c_z) are the constant elastic coefficients to be minimized using the least squares method. Then, the overall estimated forces, F_x, F_y, F_z, of the sensing device are defined as the average of the single forces:

$$[F_x, F_y, F_z] = \left[\frac{1}{N}\sum_{i=1}^{N} f_x, \frac{1}{N}\sum_{i=1}^{N} f_y, \frac{1}{N}\sum_{i=1}^{N} f_z\right], \tag{2}$$

where N indicated the total number of markers in the pattern. This approach assumes that every marker has the same impact on the estimated force and the linear relation between horizontal/vertical displacements and forces. The model has the advantage of being easy to understand, to be implemented, and can be improved by adding further relationships between the force and the geometrical properties. However, it only works for flat surfaces and cannot be directly applied to our case. We overcame this limitation by considering all the markers as independent elements and creating a stiffness vector ($N \times 3$). This can be applied to Eq. (1) and to estimate the stiffness for each marker region. We then compensated the applied force using stiffness-isolines having decreasing coefficients from the central to the peripheral area

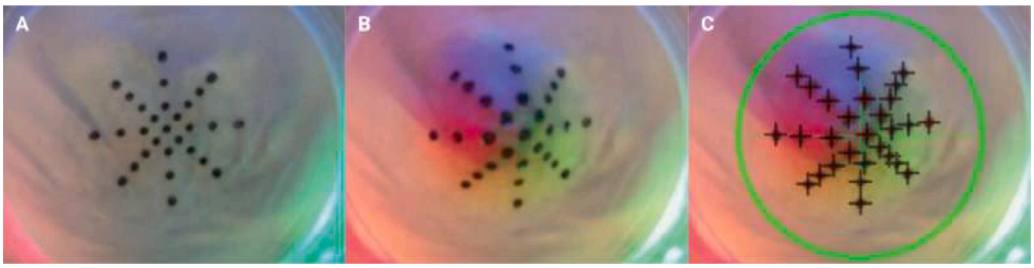

Fig. 3. Marker identification. (A) The reference frame when no forces are applied. (B) The results of the applied deformation. It is possible to notice that the markers deform according to the direction of the applied force. (C) The frame after marker detection. The green circle indicates the region of interest where the markers are supposed to be, elements outside the region are automatically removed.

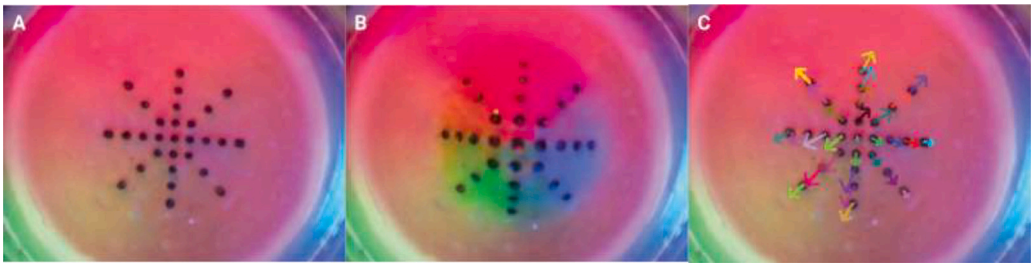

Fig. 4. Marker tracking. (A) The reference frame when no forces are applied. (B) The results of the applied deformation: 4 N applied centrally to the dome (C) The superimposed arrows show the direction of the displacement of each marker. The arrows are scaled up by a factor of 8 to make them more visible.

of the dome. After the training, the force estimation can be done either using a stiffness scalar computed as the mean value for every stiffness vector or as the resultant forces of the sum of the single force components applied to each marker.

Other than the linear approach, we implemented two machine learning models to estimate the applied forces. In the specific, we applied first a K-Neighbors Regressor (KNR) model (Yao and Ruzzo, 2006), which can be used when the labels of the dataset are continuous rather than discrete variables. Then, we considered another binary classification technique re-adapted to be used for continuous classification problems: Support Vector Regression (SVR) (Drucker et al., 1996). SVR does not natively support multi-class classification, thus, we trained 3 different models to estimate each component of the force. For both techniques, we performed a rigorous tuning of the hyperparameters to achieve the maximum performance, minimizing the mean square error (MSE) and not overfitting the model.

The last model used to estimate the applied forces was based on a Deep Convolutional Neural Network (Deep-CNN, or just CNN) (Alzubaidi et al., 2021). In particular, several models were tested, and based on the MSE metric, we chose the best-performing architecture and the respective hyperparameters: ResNet50. As for their definition, CNN is built to work with images, thus, this is the only case in which we did not use the extracted time series, but the images (raw or binarized) instead. This motivates the choice to set the LEDs of different colors.

2.4. External imaging system for fruit detection

External to the sensorized gripper, we also integrated a stereo vision system based on an RGB-D Camera (Intel Realsense D345i, Intel, USA). Such devices are widely used in fruit detection and localization as they provide depth and infrared information in addition to RGB data (Fu et al., 2020). To mount the camera a customized support has been developed and printed (X1-Carbon, BambuLab, China) using PLA. The position of the camera can be adjusted vertically to avoid having part of its field of view (FOV) covered by the sensorized units. The camera is then used to first identify the presence of a plant, then to identify and classify the status of the fruit. The first task is performed by exploiting a pre-trained CNN (YOLOV8 Jocher et al., 2023) capable of recognizing different objects and animals. Once the plant is found, the information from the camera is used to identify its location in the space by computing the barycenter of the bounding box containing the plant. A second CNN (based on YOLOV8) was used to identify the strawberries. To improve the performance, the network was fine-tuned using pre-trained weights using COCO and a freely available dataset (StrawDI_Db1) containing more than 3000 images of fully and partially visible strawberries (Pérez-Borrero et al., 2020). If a strawberry is found, the system returns its spatial position (estimated as the barycenter of the bounding box containing the fruit) and the ripe status of the fruit.

3. Results and discussion

We performed a set of experiments to first characterize the sensorized gripper and then validate its integration into the robotic system. All the experiments have been carried out in laboratory settings, trying to replicate the configurations in a hydroponic orchard. Field experiments will be considered in future works.

3.1. Force estimation

To properly use the sensor for precise picking applications, it has to be first characterized. Calibration has been performed by first acquiring the ground truth forces using an external 6-axis force/torque sensor (ATI Nano17, ATI, USA). The load cell has been connected to a motorized, vertical slider and placed above the sensing unit. After positioning and centering the load cell above the dome, while slowly moving downwards to generate pressure on the dome, we collected a set of force measurements and a set of raw frames from the internal camera. We acquired data from 9 different experiments (each configuration was acquired 5 times), changing the vertical displacement in the range between 3 mm to 12 mm with a fixed increment of 1 m. The data was then processed to first synchronize the measurements, then to extract the relevant features (e.g., timestamp, marker trajectory, marker displacement). After tracking the 29 markers, they are linearly interpolated and filtered with a Butterworth low-pass filter to reduce errors in the modeling.

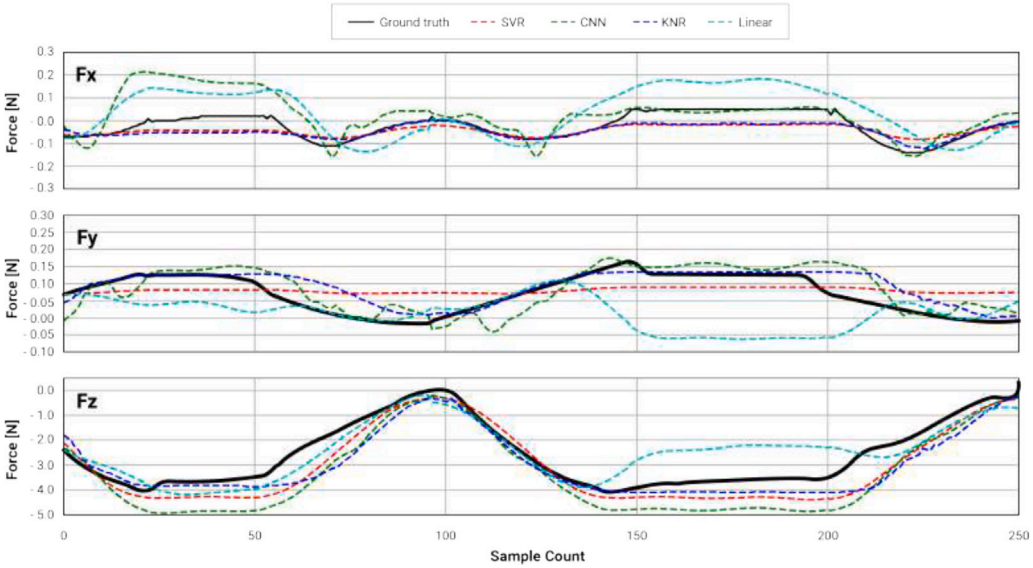

Fig. 5. Comparison of the estimated force components, considering as features the average displacement of the markers (both directions) and the average change in the radii.

Once all the raw data was cleaned and synchronized, we extracted a set of features to define the ones to be used. We implemented and tested 8 different methods for feature extraction. These are a combination of average absolute displacement, average radius increment and decrement, absolute radius value, sorted coordinates of the markers, sorted markers' radius, and sorted displacements. The dataset was then split between train and test sets with an 80–20 ratio, using a randomized approach to validate the performance of the different models. To give a quantitative measure, the training set is composed of 897 samples, following the validation set with 100 samples and the testing set with 250 samples. All the tests have been performed on a computer mounting an Intel Core i5-10400, 3 GHz, 32 GB of RAM, and without the GPU acceleration.

Among all the possible combinations, when the number of features used is limited to just the average changes both in the displacement and in the radii, the system had the worst performance. The best result was obtained with K-Neighbors Regressor (KNR) achieving the 14.1% of correct estimation. Increasing the number of features to 58—i.e., by using sorted horizontal and vertical coordinates of the markers—, the best performance was again obtained by KNR, achieving an MSE of 9.6%. Similarly, by increasing the feature to 87—i.e., by also including the displacement of the radii—, KNR outperformed all the other methods, achieving an MSE of 9.8%. By reducing the feature number to 58 and selecting a different combination—i.e., sorted horizontal and vertical displacements of the markers—, NKR obtained the best overall performance over all cases, obtaining an MSE of 2.6%. For comparison, on average SVR achieved an MSE of 35% and the compensated linear regression an MSE of 300%, respectively. The reason for the high MSE in the linear model is due to the high fluctuation of the results. It would be possible to reduce the noise by filtering the results; however, it will compromise the real-time performance of the model since it requires collecting sliding widows of samples over time, thus, increasing the processing time.

Regarding the CNN, we used the ImageNet pre-trained model with a learning rate of 0.001, a dropout probability of 0.5, and an Adam optimizer, that we trained for 50 epochs. Results show that it achieved comparable results with the other machine learning algorithms in terms of performance (average MSE 27%), with the main difference that the required time to retrieve predictions has to be taken into account. In fact, the estimations of the previous methods were obtained, on average, in 0.007 s keeping the operation rate to 20 Hz as for the frame rate. The CNN model, instead, requires 0.5 s to return the estimated force, making it difficult to be used in a real application. Fig. 5 shows the comparison of the results considering as a feature the average displacement of the markers (both directions) and the average change in the radii. For the specific case, considering only the force along the z-axis, SVR achieved an MSE value of 46.33%, CNN a value of 87.09%, KNR a value of 46.05%, and the compensated linear model a value of 53.39%.

To further prove the performance of the models, we tested them on a new set of data containing dynamic measurements obtained by applying variable forces on the domes (Fig. 6). As in the previous cases, we used as a feature the sorted coordinates and the radii of the markers. Here, SVR achieved an MSE value of 11.34%, CNN a value of 71.84%, KNR a value of 13.08%, and the compensated linear model a value of 40.0%. Differently from the training set, due to the nature of the data and the non-linearity of the model, SVR showed a better performance than KNR proving that the two models are interchangeable. Table 1 shows the best-achieved results by every force estimation approach in terms of MSE of the F_z components of the force. According to the results, KNR and SVR have similar performances both in the execution time (they can achieve real-time with the same FPS of the video streaming) and in the MSE metric. Overall KNR has a slightly better performance (0.28% lower MSE) when compared to SVR, thus, we chose it to be used during the integration test. However, the two models are interchangeable, and using one instead of the other does not change the results presented in this work.

3.2. Evaluation of the system integration

The sensorized gripper was then connected to a Franka Emika Panda robotic arm. It is a 7-DOFs robotic arm with 855 mm of reach and a maximum payload of 3 kg. The whole system is controlled at a high level by a computer running ROS 2 (Macenski et al., 2022) which acquires and processes the data from the RGB-D camera, controls the motion of the robot, and acquires and processes the data from the soft sensors. Before its use, the system needs to be registered and referenced. This allows having all the points acquired from the RGB-D camera in the same reference frame as the robot. This is performed with the classical hand–eye calibration, which returns the relative position and orientation of the camera with respect to the end-effector of the robot. At the current state, the real-time pipeline requires some supervision, which is only limited to the confirmation to proceed to the next stage of the implemented state machine.

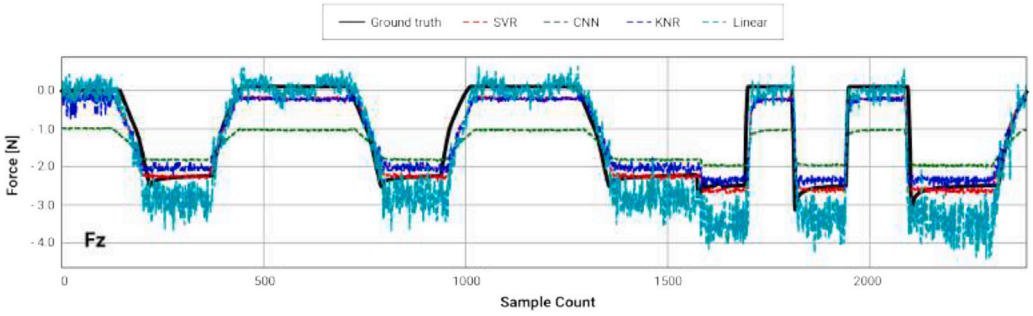

Fig. 6. Force estimation results in a dynamic case. A new set of data has been acquired and used to test the performance of the models. The estimations are compared to the ground truths acquired with the 6-axis force/torque sensor, Nano17. As results show, KNR and SVR have still similar results, but due to the nature of the data and the non-linearity of the model, SVR performs better than KNR.

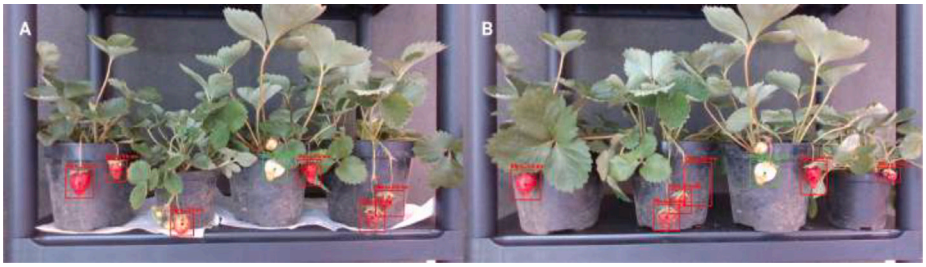

Fig. 7. Results of the CNN for fruit ripeness identification. (A) A set of all positive results in which the network correctly identifies the ripen and unripe fruits. In the picture, two fruits were not correctly identified. However, the fruits are completely unripe, so not to be listed among the possible targets. (B) An error in the identification of a possible target: a part of the jar is classified as a ripe strawberry. In this case, further controls should be introduced to avoid a wrong picking.

Table 1
Summary of the best performance achieved by every model evaluated on the test set (MSE values refer to the F_z component only).

Model	Number of features	Feature scaling	Best MSE
Linear regression with compensation	58		≥200%
K-Neighbors Regressor	58	No	2.86%
Support vector regression	87	Yes	11.36%
Deep Convolutional Neural Network	224 × 224 × 3		14.38%

To demonstrate the integrated system capabilities, we recreated a scenario where a strawberry plant is positioned in a classical hydroponic orchard (Fig. 1C). The robotic arm, which is fixed to a mobile platform, was placed at a distance of 500 mm from the plants to ensure the reachability of the target and to avoid complex configurations that can bring to singularities. Once activated, the system starts looking for the plants using the RGB-D camera and the pre-trained network. The main advantage of using such an approach is the speed in the recognition of a set of common objects. However, even if the correct identification rate is higher than 98.2%, the bounding box created around a target does not always contain a single object (i.e., a single strawberry) or the expected result (i.e., the network provides a false positive). This error is frequent when multiple objects of the same type are placed in proximity, which is a common case for hydroponic orchards. Nevertheless, the issue can be solved by skipping this initial phase and assuming to know the location of the plant and just focusing on the fruits alone. Fig. 7 shows a set of classification results provided by the CNN.

Once the position of the plant is obtained, by using the geometric center of the detected bounding box, the robot starts to approach the target while looking for the presence of strawberries. The fruit-detection network has a lower success rate of 92.4%. This was mostly due to the position of the strawberries on the plant (i.e., close to the center of the pot and not hanging out of it) which were most prone to be covered with leaves. Fig. 8 depicts a schematic representation of the sequential process used to detect and collect the strawberries.

According to the different types of targets, the robot applies different approaching strategies. If the target is the plant, it plans and executes a linear trajectory to approach a centered target Cartesian point, so that the camera can have a closer view of the target. Contrary, if the target is the strawberry, the linear trajectory brings the gripper to a suitable position centered with respect to the fruit so that picking can be performed. All the trajectories are controlled in position and follow a set of waypoints computed with a sampling frequency of 1 kHz. For safety reasons, we reduced by 25% the maximum speed of the robot and forced the trajectories to be performed in a fixed amount of time, 5 s. We evaluated the performance of the robot controller and the vision system by repeating the approaching tasks for a total amount of 100 repetitions, varying the starting configurations and the position of the plants/fruits. The system can correctly position the gripper with a maximum error of 10 mm computed from the reference frame of the gripper (Fig. 2). The issue does not compromise the system performance since it can be repositioned accordingly to the computed contact point. During the experiments, repositioning occurred 10% of the time, and the correct picking after the repositioning was 60%.

When the robot reaches the final target (i.e., the strawberry to be picked), the system activates the gripper, which closes until a first contact is measured. Then, its closure is controlled in real-time using a force feedback (i.e., close-loop controller based on the developed sensor) until reaching the empirical threshold of 1.75 N. Afterward, the strawberry is harvested from the plant by moving the gripper back, while slightly rotating it toward the ground imitating the common picking used in manual harvesting by humans. The whole picking task is performed in less than 10 seconds (from the start of the grasping to detach of the fruit). Due to the limited availability of plants with fruits

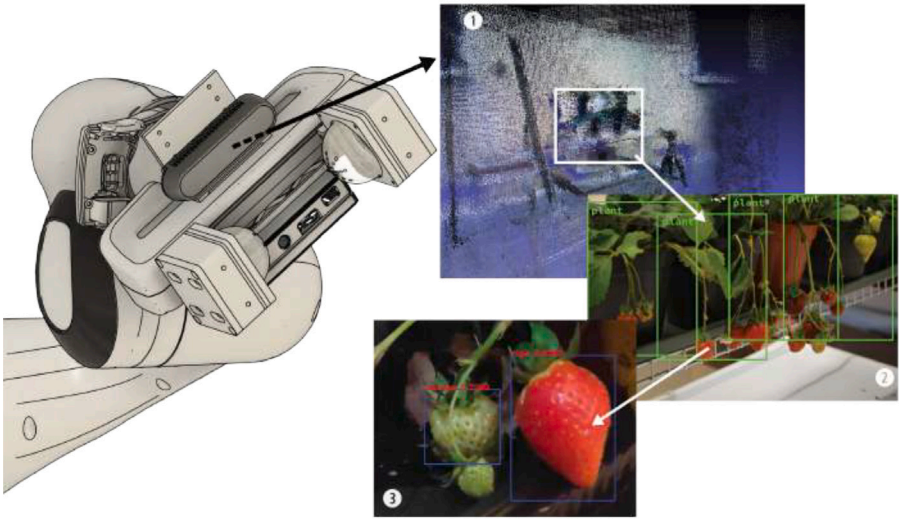

Fig. 8. A schematic representation of the workflow performed by the external vision system and the robot. (1) The robot reconstructs a 3D map of the environment while looking at plants. (2) When identified, each plant is considered independently and the robot, while approaching it, starts to search for strawberries. (3) If strawberries is found, the system identifies the state of ripe, and, if ready to be picked, commands the robot to move toward them one-by-one.

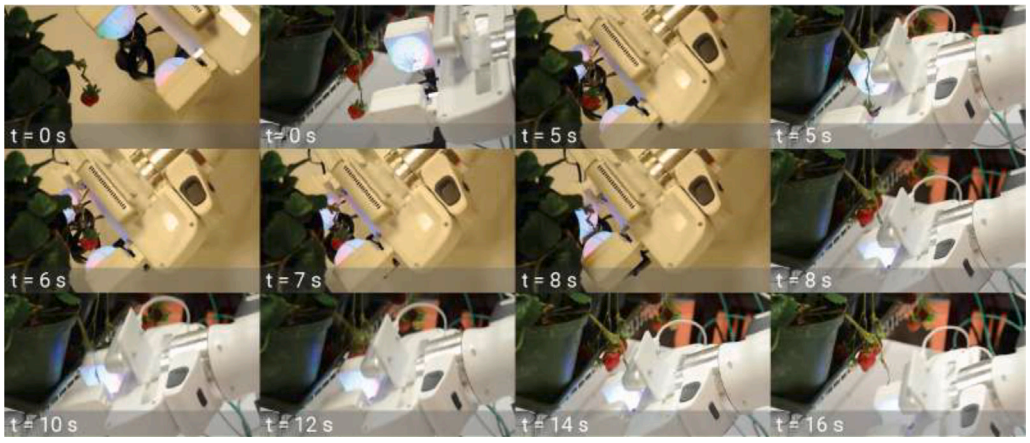

Fig. 9. Selected frame sequence of a picking task; the frames are time-coded to facilitate the sequencing. To better see the different phases and the motion of the gripper and the robot, the frames are taken from two angles.

(experiments were carried out in winter), we tested the system in a period of 6 months on 28 plants, and on a total 50 fruits with a diameter in the range of 15 to 48 mm. With this method, we successfully harvest, with a single attempt, the 82% of the correctly identified strawberries without damaging the fruit. Fig. 9 shows a sequence of selected frames taken from two angles during the picking of a fruit.

4. Conclusions

In this paper, we present the integration of a soft, sensorized gripper with a robotic system able to identify and pick small fruits such as strawberries. Soft grippers can be considered as one of the best solutions for harvesting, thanks to their adaptability and delicacy when grasping and manipulating the target products. In fact, by using materials with a module of elasticity similar to biological materials, soft grippers ensure safe interaction with humans and the working environment.

The gripper consists of a patterned, hemispherical dome that is placed above a camera. By tracking the positions and the radii of the markers in the pattern upon external deformations, and applying machine learning to the data, it was possible to estimate the applied force. Compared to other similar solutions, which focus on the estimation of slip and normal forces, in this work we integrated the soft tactile sensor into a robotic system that comprises an anthropomorphic arm (Franka Emika Panda) and a gripper-mounted vision system to detect the plants and each of the fruits. The system can correctly identify the targets with an accuracy of 92% and can position the gripper with a maximum error of 10 mm computed from the reference frame of the gripper (Fig. 2). Using this method, we could harvest 82% of the identified targets without damaging any fruit or leaving the stalk still attached to it. The latter feature is an improvement to the solution already available which requires subsequent processing to remove the stalk to avoid to damage strawberries when they are placed in baskets for commercialization.

The proposed force estimation techniques rely on computer vision, machine learning algorithms, and the calibration of the device. This means that software updates could surely improve the estimation performance and add new features without necessarily changing the design of the gripper. In addition to further analysis in real-world scenarios, future developments could focus on this subject, by providing a fully automated calibration process.

In the agri-robotic domain, execution time is a critical aspect, especially in extensive farms. In this work, we just focused on a static condition where the robot is already in position in front of a set of plants and does not need to move between different sections of the farm. Limiting with this specific case, and considering the constraints

introduced earlier, the performance of the robot is comparable with other fully integrated robotic systems (De Preter et al., 2018; Xiong et al., 2020) that require an average of 5 s to harvest fruit with multiple attempts (success rate from 40 to 70%). As an additional future work, the picking rate can be improved by considering different directions to approach the fruits and not only moving horizontally toward the targets. All of these should be carried out to increase the harvest ratio between robotic and manual picking, increase the harvested fruit quality, and make robotic solutions economically justified to put into practice the Industry 4.0 transition also for the agricultural sector.

CRediT authorship contribution statement

Francesco Visentin: Conceptualization, Methodology, Validation, Investigation, Formal analysis, Writing – original draft, Writing – review & editing, Visualization. **Fabio Castellini:** Software, Validation, Investigation, Data curation. **Riccardo Muradore:** Writing – review & editing, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

Ahmed, F., Waqas, M., Jawed, B., Soomro, A.M., Kumar, S., Hina, A., Khan, U., Kim, K.H., Choi, K.H., 2022. Decade of bio-inspired soft robots: a review. Smart Mater. Struct. 31 (7), 073002.

Alspach, A., Hashimoto, K., Kuppuswamy, N., Tedrake, R., 2019. Soft-bubble: A highly compliant dense geometry tactile sensor for robot manipulation. IEEE.

Alzubaidi, L., Zhang, J., Humaidi, A.J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M.A., Al-Amidie, M., Farhan, L., 2021. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. J. Big Data 8 (1), 53.

Amend, J., Cheng, N., Fakhouri, S., Culley, B., 2016. Soft robotics commercialization: Jamming grippers from research to product. Soft Robot. 3 (4), 213–222.

Anon, 2019. Development and field evaluation of a strawberry harvesting robot with a cable-driven gripper. Comput. Electron. Agric. 157, 392–402.

Anon, 2023. A flexible swallowing gripper for harvesting apples and its grasping force sensing model. Comput. Electron. Agric. 204, 107489.

Arima, S., Kondo, N., Monta, M., 2004. Strawberry harvesting robot on table-top culture. In: 2004 ASAE Annual Meeting. American Society of Agricultural and Biological Engineers, p. 1.

Blanes, C., Mellado, M., Beltran, P., 2014. Novel additive manufacturing pneumatic actuators and mechanisms for food handling grippers. Actuators 3 (3), 205–225.

Brown, E., Rodenberg, N., Amend, J., Mozeika, A., Steltz, E., Zakin, M.R., Lipson, H., Jaeger, H.M., 2010. Universal robotic gripper based on the jamming of granular material. Proc. Natl. Acad. Sci. 107 (44), 18809–18814.

Chen, W., Khamis, H., Birznieks, I., Lepora, N.F., Redmond, S.J., 2018. Tactile sensors for friction estimation and incipient slip detection—Toward dexterous robotic manipulation: A review. IEEE Sens. J. 18 (22), 9049–9064.

Choi, S.-H., Tahara, K., 2020. Dexterous object manipulation by a multi-fingered robotic hand with visual-tactile fingertip sensors. ROBOMECH J. 7 (1).

De Preter, A., Anthonis, J., De Baerdemaeker, J., 2018. Development of a robot for harvesting strawberries. IFAC-PapersOnLine 51 (17), 14–19, 6th IFAC Conference on Bio-Robotics BIOROBOTICS 2018.

Do, W.K., Kennedy, M., 2022. DenseTact: Optical tactile sensor for dense shape reconstruction. In: ICRA.

Drucker, H., Burges, C.J.C., Kaufman, L., Smola, A., Vapnik, V., 1996. Support vector regression machines. In: Mozer, M., Jordan, M., Petsche, T. (Eds.), Advances in Neural Information Processing Systems. MIT Press.

Elfferich, J.F., Dodou, D., Santina, C.D., 2022. Soft robotic grippers for crop handling or harvesting: A review. IEEE Access 10, 75428–75443.

Feng, Q., Wang, X., Zheng, W., Qiu, Q., Jiang, K., 2012. New strawberry harvesting robot for elevated-trough culture. Int. J. Agric. Biol. Eng. 5 (2), 1–8.

Fernandez, A.J., Weng, H., Umbanhowar, P.B., Lynch, K.M., 2021. Visiflex: A low-cost compliant tactile fingertip for force, torque, and contact sensing. IEEE.

Fitzgibbon, A., Pilu, M., Fisher, R., 1999. Direct least square fitting of ellipses. IEEE Trans. Pattern Anal. Mach. Intell. 21 (5), 476–480. http://dx.doi.org/10.1109/34.765658.

Fu, L., Gao, F., Wu, J., Li, R., Karkee, M., Zhang, Q., 2020. Application of consumer RGB-D cameras for fruit detection and localization in field: A critical review. Comput. Electron. Agric. 177, 105687.

Galley, A., Knopf, G.K., Kashkoush, M., 2019. Pneumatic hyperelastic actuators for grasping curved organic objects. Actuators 8 (4).

Hughes, J., Culha, U., Giardina, F., Guenther, F., Rosendo, A., Iida, F., 2016. Soft manipulators and grippers: A review. Front. Robot. AI 3.

Ilievski, F., Mazzeo, A.D., Shepherd, R.F., Chen, X., Whitesides, G.M., 2021. Soft robotics for chemists. Angew. Chem. Int. Ed. 50 (8), 1890–1895.

Jocher, G., Chaurasia, A., Qiu, J., 2023. Ultralytics YOLOv8. URL https://github.com/ultralytics/ultralytics.

King, J.P., Bauer, D., Schlagenhauf, C., Chang, K.-H., Moro, D., Pollard, N., Coros, S., 2018. Design. Fabrication, and evaluation of tendon-driven multi-fingered foam hands. In: 2018 IEEE-RAS 18th International Conference on Humanoid Robots. Humanoids, IEEE, pp. 1–9.

Kuppuswamy, N., Alspach, A., Uttamchandani, A., Creasey, S., Ikeda, T., Tedrake, R., 2020. Soft-bubble grippers for robust and perceptive manipulation. Soft Robot.

Kuriyama, Y., Okino, Y., Wang, Z., Hirai, S., 2019. A wrapping gripper for packaging chopped and granular food materials. In: 2019 2nd IEEE International Conference on Soft Robotics. RoboSoft, pp. 114–119.

Liu, C.-H., Chiu, C.-H., Chen, T.-L., Pai, T.-Y., Chen, Y., Hsu, M.-C., 2018. A soft robotic gripper module with 3D printed compliant fingers for grasping fruits. In: 2018 IEEE/ASME International Conference on Advanced Intelligent Mechatronics. AIM, pp. 736–741.

Lowenberg-DeBoer, J., Huang, I.Y., Grigoriadis, V., Blackmore, S., 2020. Economics of robots and automation in field crop production. Precis. Agric. 21 (2), 278–299.

Macenski, S., Foote, T., Gerkey, B., Lalancette, C., Woodall, W., 2022. Robot operating system 2: Design, architecture, and uses in the wild. Science Robotics 7 (66), eabm6074.

Mizushima, K., Oku, T., Suzuki, Y., Tsuji, T., Watanabe, T., 2018. Multi-fingered robotic hand based on hybrid mechanism of tendon-driven and jamming transition. In: 2018 IEEE International Conference on Soft Robotics. RoboSoft, IEEE, pp. 376–381.

Navas, E., Fernández, R., Sepúlveda, D., Armada, M., de Santos, P.G., 2021. Soft grippers for automatic crop harvesting: A review. MDPI.

Ozawa, R., Tahara, K., 2017. Grasp and dexterous manipulation of multi-fingered robotic hands: a review from a control view point. Adv. Robot. 31 (19–20), 1030–1050.

Pérez-Borrero, I., Marín-Santos, D., Gegúndez-Arias, M.E., Cortés-Ancos, E., 2020. A fast and accurate deep learning method for strawberry instance segmentation. Comput. Electron. Agric. 178, 105736.

Sakuma, T., Phillips, E., Ricardez, G.A.G., Ding, M., Takamatsu, J., Ogasawara, T., 2019. A parallel gripper with a universal fingertip device using optical sensing and jamming transition for maintaining stable grasps. In: 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems. IROS, pp. 5814–5819.

Sakuma, T., von Drigalski, F., Ding, M., Takamatsu, J., Ogasawara, T., 2018. A universal gripper using optical sensing to acquire tactile information and membrane deformation. IEEE.

Scharff, R.B., Boonstra, D.-J., Willemet, L., Lin, X., Wiertlewski, M., 2022. Rapid manufacturing of color-based hemispherical soft tactile fingertips. In: RoboSoft 2022.

Tawk, C., Gillett, A., in het Panhuis, M., Spinks, G.M., Alici, G., 2019. A 3D-printed omni-purpose soft gripper. IEEE Trans. Robot. 35 (5), 1268–1275.

Visentin, F., Murali Babu, S.P., Meder, F., Mazzolai, B., 2021. Selective stiffening in soft actuators by triggered phase transition of hydrogel-filled elastomers. Adv. Funct. Mater. 31 (32), 2101121.

Vulliez, P., Gazeau, J.-P., Laguillaumie, P., Mnyusiwalla, H., Seguin, P., 2018. Focus on the mechatronics design of a new dexterous robotic hand for inside hand manipulation. Robotica 36 (8), 1206–1224.

Wang, Z., Hirai, S., 2018. Chamber dimension optimization of a bellow-type soft actuator for food material handling. In: 2018 IEEE International Conference on Soft Robotics. RoboSoft, pp. 382–387.

Wang, Z., Kanegae, R., Hirai, S., 2021. Circular shell gripper for handling food products. Soft Robot. 8 (5), 542–554.

Wang, Z., Or, K., Hirai, S., 2020. A dual-mode soft gripper for food packaging. Robot. Auton. Syst. 125, 103427.

Xiong, Y., From, P.J., Isler, V., 2018. Design and evaluation of a novel cable-driven gripper with perception capabilities for strawberry picking robots. In: 2018 IEEE International Conference on Robotics and Automation. ICRA, pp. 7384–7391.

Xiong, Y., Ge, Y., Grimstad, L., From, P.J., 2020. An autonomous strawberry-harvesting robot: Design, development, integration, and field evaluation. J. Field Robotics 37 (2), 202–224.

Yamaguchi, A., 2018. FingerVision for tactile behaviors, manipulation, and haptic feedback teleoperation. IEEJ.

Yamaguchi, A., Atkeson, C.G., 2017. Implementing tactile behaviors using FingerVision. IEEE.

Yao, Z., Ruzzo, W.L., 2006. A regression-based k nearest neighbor algorithm for gene function prediction from heterogeneous data. BMC Bioinformatics 7 (1), S11.

Zhang, Y., Kan, Z., Yang, Y., Tse, Y.A., Wang, M.Y., 2019. Effective estimation of contact force and torque for vision-based tactile sensors with Helmholtz–Hodge decomposition. IEEE Robot. Autom. Lett. 4 (4), 4094–4101.

Zhang, B., Xie, Y., Zhou, J., Wang, K., Zhang, Z., 2020. State-of-the-art robotic grippers, grasping and control strategies, as well as their applications in agricultural robots: A review. Comput. Electron. Agric. 177, 105694.