130. An active laser-camera scanning system for precision fruit localization in robotic harvesting

K. Zhang[1], P. Chu[2], K. Lammers[2], Z. Li[1] and R. Lu[3]
[1]*Department of Mechanical Engineering, Michigan State University, East Lansing, MI, USA*
[2]*Department of Electrical and Computer Engineering, Michigan State University, East Lansing, MI, USA*
[3]*U.S. Department of Agriculture Agricultural Research Service (USDA-ARS), East Lansing, MI, USA;*
** renfu.lu@usda.gov*

Abstract

An active laser-camera scanner (ALACS), mainly consisting of a red line laser, an RGB camera, and a linear motion slide, was designed for robust and precision fruit localization. A dynamic-targeting laser-triangulation method, coupled with a laser line extraction algorithm, was developed for precise depth computation of target fruits. Calibration results showed that ALACS had the maximum localization errors of 0.6 mm, 1.2 mm and 4.0 mm in the *x*, *y* and *z* (depth) directions, respectively, and the average depth error of less than 1 mm for targets at distances between 100 cm and 160 cm. When integrated with a vacuum-based robotic apple harvesting system, ALACS achieved a 95% fruit detachment rate in an orchard picking evaluation.

Keywords: apples, fruit detection, fruit localization, laser triangulation, robotic harvesting

Introduction

Fruit detection and localization is a critical task in robotic harvesting. Due to variable natural lighting conditions, colour variations of fruits, and fruit occlusions by foliage and branches, robust and effective fruit detection and localization in the real orchard environment poses significant technical challenges. Fruit detection is commonly accomplished using colour images captured by RGB cameras. The acquired colour images are processed and segmented for identifying targe fruits, using either feature-based or deep learning-based methods. In recent years, deep learning-based object recognition algorithms have been increasingly used for fruit detection with promising results (Bargoti and Underwood, 2017; Chu *et al*., 2021). After fruits are detected, their spatial positions need to be determined, so that the robot manipulator can reach to the target fruits for picking. Different sensing techniques have been used for fruit localization (Gené-Mola *et al*., 2019; Kang and Chen, 2020). One example is (passive) stereo vision systems, which exploit the two-camera layout and triangulation optical measurement principle to obtain depth information (Lazaros *et al*., 2008). Stereo vision systems are expensive in computation and the performance of stereo matching is affected by occluded pixels or varying lighting conditions that are common in the orchard environment. RGB-depth (or RGB-D) cameras are another type of depth measurement sensors to localize fruits (Xiong *et al*., 2019). Different from passive stereo vision systems that solely rely on natural light, the RGB-D sensors include a separate artificial illumination source to aid the depth computation. According to the methods on how depth measurements are computed, RGB-D cameras can be further divided into structured light (SL), time of flight (ToF), and active infrared stereo (AIRS) (Fu *et al*., 2020). Despite some successes, these RGB-D cameras still have limited and unstable performance in the natural orchard environment. For example, the SL-based sensors are sensitive to the natural light condition and the interference of multiple patterned light sources. ToF systems are vulnerable to scattered light and multi-path interference, and have lower depth resolutions, compared to other RGB-D cameras. Similar to passive stereo vision systems, AIRS-based sensors also encounter stereo matching issues, which can lead to flying pixels or over-smoothing around the contour edges (Fu *et*

al., 2020). In addition, the performance of these sensors could deteriorate significantly when target fruits are occluded by leaves and branches, due to low or limited density of the illuminating light patterns or point cloud.

In view of the shortcomings of the existing 3D imaging systems, a new 3D localization technique was proposed, based on the laser triangulation principle, to improve fruit localization accuracy and robustness for robotic harvesting of apples. This paper reports on the design, calibration, fruit localization, and performance evaluation of an active laser-camera scanning (ALACS) system.

Materials and methods

Setup of the ALACS

The ALACS unit is mainly comprised of a red line laser, a FLIR RGB camera (Model BFLY-U3-23S6C-C) with 1920×1200 pixels and 70°×48° (V×H) field of view, and a linear motion slide (Figure 1). The linear motion slide enables the laser to move horizontally. Meanwhile, the RGB camera is installed at one (right) end of the linear motion slide with an angle relative to the laser. The ALACS obtains depth measurements from a target, based on the laser triangulation principle, by pairing the laser illumination source with the camera, which is widely used for precision 3D object profiling. Unlike conventional laser triangulation sensors, in which the relative position between the laser and the camera is fixed, with the ALACS, the camera is fixed while the laser moves with the linear motion slide. A continuous-wave red line laser at 635 nm was selected, because this wavelength provides stronger contrast on red-coloured apples, thus facilitating the extraction of laser lines or line segments from the image. An RGB-D camera (not shown in Figure 1) and the ALACS are fused synergistically to achieve apple detection and localization. Specifically, the fusion scheme includes two steps. In the first step, the RGB-D camera captures an image from the workspace, which is then fed into a deep learning-based image segmentation algorithm for fruit detection (Chu *et al*., 2023), from which an initial rough estimation of the target apple location is obtained. In the second step, by using the rough apple location information provided by the RGB-D camera, the ALACS is triggered to scan the target apple to obtain an improved estimation of the fruit location. As shown in Figure 1, the basic working principle of the ALACS is to project the laser line onto the target fruit and then use the image information and triangulation technique to localize the fruit. The perception strategy of ALACS is designed as follows:

1. Initialization. The linear motion slide is actuated to move the laser towards an initial position, so that the red laser line is approximately projected on the left-half region of the target apple. The initial laser position is obtained by transforming the rough target apple location provided by the RGB-D camera into the coordinate frame of the ALACS camera.
2. Interval scanning. After the laser has reached the initial position, the camera is activated to capture an image. The linear motion slide then travels to the right in 1-cm increments for four more times. This scanning strategy would mitigate the impact of leave occlusion, since the laser line provides high spatial-resolution localization information for the target fruit and reduce the likelihood of the laser lines being entirely blocked by obstacles such as leaves and branches.
3. Refinement of 3D position. For each image captured by the camera, the laser line projected on the target apple surface is extracted and then used to generate 3D location information. Image processing and laser triangulation techniques are applied to accomplish laser line extractions and 3D position computations. Finally, a holistic evaluation function is used to select one of the candidate laser lines or line segments to be the target apple location.

To accomplish the aforementioned fruit localization scheme, laser line extraction and position candidate computation are needed. The laser line extraction is achieved by leveraging computer vision techniques (Chu *et al*., 2024). To facilitate the computation of fruit 3D positions, a high-fidelity model was derived based on the principle of laser triangulation and a robust calibration scheme.

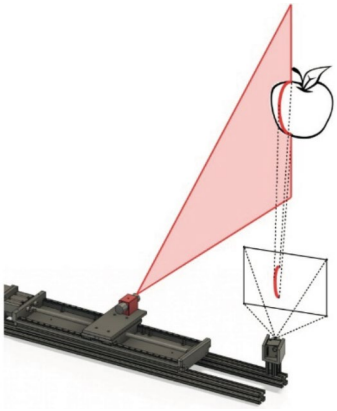

Figure 1. The working principle of the active laser-camera scanning system.

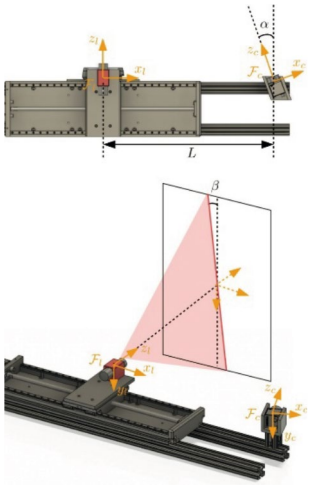

Figure 2. Coordinate frames and extrinsic parameters of the ALACS unit.

Laser triangulation

The basic idea of laser triangulation-based technique is to capture depth measurements by pairing a laser illumination source with a camera. Both the laser beam and the camera are aimed at the target object and based on the extrinsic parameters between the laser source and the camera sensor, the depth information can be calculated by using trigonometry. As shown in Figure 2, $\mathcal{F}_l$ and $\mathcal{F}_c$ are denoted as the laser frame and camera frame, respectively. α is the rotating angle along the y_l-axis between $\mathcal{F}_l$ and $\mathcal{F}_c$. L is the horizontal distance (i.e., the translation along the x_l-axis) between $\mathcal{F}_l$ and $\mathcal{F}_c$. β is the angle between the laser plane and the (y_l, z_l) plane of $\mathcal{F}_l$. α, L, and β are the extrinsic parameters between the laser source and the camera and used for deriving the high-fidelity model of the ALACS. Based on the pin-hole model of the camera, without going into details (Zhang *et al.*, 2024), the following localization model for ALACS can be obtained:

$$x_{c,i} = \frac{L\bar{u}_{c,i}}{\sin(\alpha) - \bar{u}_{c,i}\cos(\alpha) - \bar{v}_{c,i}\tan(\beta)} \tag{1}$$

$$y_{c,i} = \frac{L\bar{v}_{c,i}}{\sin(\alpha) - \bar{u}_{c,i}\cos(\alpha) - \bar{v}_{c,i}\tan(\beta)} \tag{2}$$

$$z_{c,i} = \frac{L}{\sin(\alpha) - \bar{u}_{c,i}\cos(\alpha) - \bar{v}_{c,i}\tan(\beta)} \tag{3}$$

where $\overline{u}_{c,i}$ and $\overline{v}_{c,i}$ are the pixel coordinates on the image plane, while α, L, and β represent the extrinsic parameters for ALACS (Figure 2). Hence, with the known pixel coordinates $\overline{u}_{c,i}$ and $\overline{v}_{c,i}$, the 3D position $p_{c,i} = [x_{c,i}, y_{c,i}, z_{c,i}]^T$ can be readily calculated from equations (1)–(3).

Robust calibration scheme

The extrinsic parameters α, L, and β play a crucial role in facilitating the 3D measurement of ALACS and require robust calibrations. Note that α and β are constants, while L is variable as the linear motion slide can move to different positions. During the calibration procedure, the linear motion slide is fixed at an initial position, and the corresponding horizontal distance between the laser and camera is denoted by L_0. α, β and L_0 (i.e., the initial value of L) are obtained via offline calibration. Then, when the linear motion slide is moving, L could be updated online based on its initial value L_0 and the movement distance of the linear motion slide.

The calibration procedure includes two steps. In the first step, multiple sets of data $s_i = [\overline{u}_{c,i}, \overline{v}_{c,i}, z_{c,i}]^T$ ($i = 1, 2, \cdots, n$) are collected from recorded images. The second step then formulates an optimization problem by using the collected data and equations (1)–(3) to

compute the extrinsic parameters. In general, the data samples $s_i = [\overline{u}_{c,i}, \overline{v}_{c,i}, z_{c,i}]^T$ ($i = 1, 2, \cdots, n$) are corrupted with noises and may contain outliers that do not satisfy the relation (1). These outliers can severely influence the calibration accuracy and thus need be removed. Towards that end, the random sample consensus (RANSAC) methodology (Raguram *et al.*, 2013) is adopted to extract credible data. The developed calibration scheme leverages RANSAC techniques to iteratively estimate the model parameters and select the solution with the largest number of inliers. Therefore, it is able to robustly identify the model parameters when some data samples are corrupted or noisy.

Laser line extraction (LLE)

After the RGB-D camera has captured a global view of the scene, both colour and depth information will be obtained and then used to segment and identify bounding boxes containing apples based on an apple detection algorithm (Chu *et al.*, 2023). Rough 3D positions of all detected apples are then estimated. Based on a planning strategy (Zhang *et al.*, 2021), a target apple is selected and its rough 3D position estimated by the RGB-D camera is then used to guide the laser to the initial position for the target apple. The laser is then moved horizontally from the left to the right direction at 1-cm increments, which result in five color images (Figure 3). These images are then processed to extract the laser lines from the surface of the apple. LLE involves four steps: laser pattern detection (from the red channel image), noise removal, line focus, and curve fitting (Figure 4). The red channel of the RGB image is chosen for laser line extraction, because preliminary studies found that this channel provides strong contrast between the laser line and the fruit. This results in five laser projections covering the target apple. The most reliable candidate is then selected to represent the centroid position of the apple based on a confidence evaluation for each candidate, which is calculated using two key factors: the distance to the estimated centre and the number of extracted laser line pixels. The distance factor in the confidence calculation quantifies how close the candidate laser line is to the apple's estimated centre. Candidates that are closer to the centre are considered more reliable and are assigned higher confidence scores. This distance factor Δd is calculated using the Euclidean distance between the candidate's position and the estimated centre. The second factor contributing to the confidence calculation is the number of extracted laser line pixels, N, for each candidate. Candidates with more extracted laser line pixels are considered to provide a more complete representation of the apple's surface geometry and are thus deemed more reliable. Consequently, these candidates are assigned higher confidence scores. Then, the confidence P is calculated by $P=\omega_1 \cdot \Delta d - \omega_2 \cdot N$, where ω_1 and ω_2 are weights for these two factors and are obtained through cross validation. After selection of the most reliable candidate based on the calculated confidence scores, the 2D position of the centre of the laser line is obtained and the apple's centre position is determined by transforming the 2D position to 3D position, using equations (1)–(3).

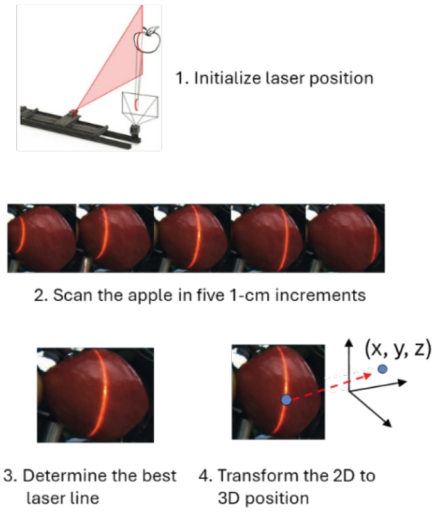

Figure 3. Procedures for acquiring laser lines from, and determining the spatial position of, a target fruit.

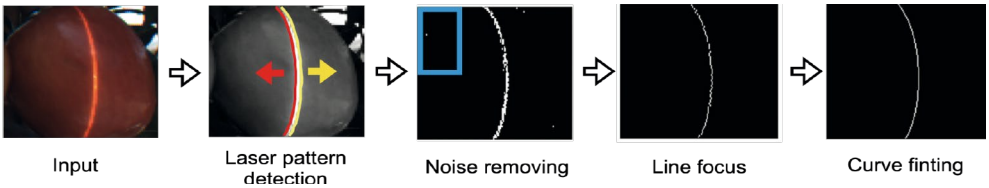

Figure 4. The workflow of extracting a laser line from the apple.

Calibrations of ALACS

Calibration of the ALACS was conducted in laboratory using a planar checkerboard. Calibration of the ALACS was conducted in laboratory using a planar checkerboard. The planar checkerboard was placed at 10 different positions in sequence, and at each position the colour camera was triggered to capture an image. For each image, three laser points were selected and the corresponding data samples $s_i = [\overline{u}_{c,i}, \overline{v}_{c,i}, z_{c,i}]^T$ were computed, using the method described earlier. A total of 30 data samples were collected and then used for the calibration of the extrinsic parameters. The high-fidelity model with RANSAC techniques was used for calibration. The mean error of $|z_{c,i} - \hat{z}_{c,i}|$ was computed to evaluate the performance of the ALACS.

The localization accuracy of the ALACS was evaluated for five different distances at 5 cm increments between the laser and the camera (i.e., $d = 0, 5, 10, 15, 20$, where $d = 0$ is the initial laser position). Given L_0 and d, L can be computed by $L = L_0 - d$. For each laser position (i.e., for each L value), 10 images were collected with the planar checkerboard being placed at different positions. Three laser points were randomly chosen from each image, and then at each laser position, a total of 30 data samples were utilized to evaluate the localization accuracy of the ALACS. 3D measurements of the collected data, i.e., $p_{c,j} = [x_{c,j}, y_{c,j}, z_{c,j}]^T$ ($j = 1, 2, \cdots, 30$), were obtained with the aid of the checkerboard setup. Meanwhile, the extrinsic parameters calculated with the developed robust calibration scheme were used to determine the estimated 3D measurements $\hat{p}_{c,j} = [\hat{x}_{c,j}, \hat{y}_{c,j}, \hat{z}_{c,j}]^T$.

Field evaluation of ALACS

After laboratory calibration and evaluation, the ALACS was integrated with our robotic apple harvesting system (Figure 5) (Zhang *et al.*, 2024). The robotic system consists of a perception

module, a 4 degree-of-freedom manipulator, a soft vacuum-based end-effector, and a dropping module. The perception module is mainly composed of an RGB-D camera and ALACS. The entire software is fully integrated using the robot operating system, where different software components are primarily communicated via custom messages.

ALACS was evaluated in a research orchard of Michigan State University in East Lansing, Michigan. In the field evaluation, the vacuum system was turned off. After the ALACS determined the position of a target fruit, the robot arm was actuated to reach to the target fruit. The distance between the target fruit and the end effector's final position was manually measured. Likewise, a commercial RGB-D camera (RealSense D435i, Intel), which was also installed with the robotic harvesting system, was evaluated for its performance in localizing the same apples that had been tested by ALACS. The vacuum-based end effector was able to pick the fruit when it was within 20 mm distance. Hence, a measured localization error larger than 20 mm would be considered a failed detachment. A total of 100 'Gala' apples were randomly chosen from trees for measurement. These apples had different degrees of occlusion by leaves; however, the degrees of fruit occlusion were not quantified in the study.

Results and discussion

Figure 6 shows the localization error distributions of ALACS with the laser being placed at five different positions. The ALACS achieved accurate localization in the *x* (horizontal), *y* (vertical), and *z* (depth) directions. The largest localization errors in the *x*, *y*, and *z* directions are less than 0.6 mm, 1.2 mm, and 4 mm, respectively. The mean depth () error for all localization measurements is only 0.39 mm. Overall, the localization errors in the three directions increase with the decreasing

Figure 5. A photograph (a) and 3-dimensional sketch (b) of the robotic apple harvesting system integrated with the active laser-camera scanning (ALACS) unit.

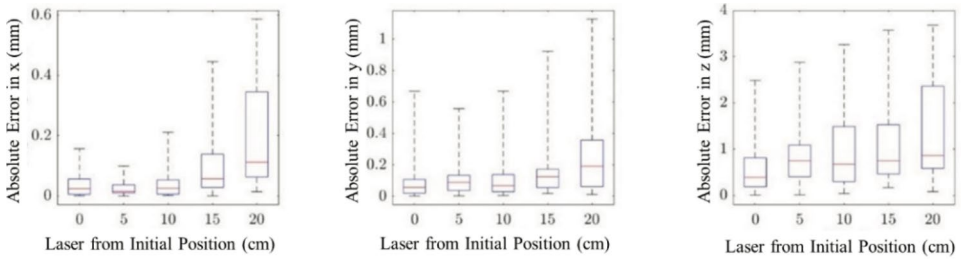

Figure 6. Localization accuracy for the laser moving from the initial position (38.2 cm to the camera) towards the camera at 5 cm increments. In each box, the red line mark is the median error, the edges of the box are the 25th and 75th percentiles, and the whiskers represent the range of the data points.

distance between the laser and camera. These results indicate that ALACS provides high accurate localization measurements for targets.

The field evaluation showed that with ALACS, the robotic harvesting system would achieve 95% fruit detachment rate versus 71% for the commercial RGB-D camera. It should be mentioned that these detachment rates were calculated based on the assumption that the robot would pick target fruits within 20 mm to the end effector. The actual detachment rate could be lower, due to such factors as fruit orientation on trees and vacuum suction performance. Overall, ALACS has shown superior performance in localizing apples on trees. Compared to other commercial 3D sensing systems, ALACS would be less influenced by the natural lighting variation, due to the use of a red laser and higher pixel density. However, the current configuration needs to scan each target fruit multiple times in sequence, which could have a negative effect on the robotic harvesting speed. Hence, further improvements to the ALACS design and scanning process are needed to achieve faster and more robust localization of apples.

Conclusions

A new active laser-camera scanning (ALACS) technique was developed for improving fruit localization accuracy and mitigating the effect of natural lighting and leave occlusion in robotic harvesting. With the robust calibration scheme, ALACS has achieved superior localization accuracies with the average errors in the three spatial directions being less than 1 mm and the maximum error being less than 4 mm when an object is at distances between 100 and 160 cm. When integrated with a vacuum-based apple harvesting system, ALACS was able to achieve 95% fruit detachment rate. Future improvements to the ALACS are needed for faster and more robust scanning of target fruits to meet the robotic fruit harvest requirement.

References

Bargoti, S., & Underwood, J.P. (2017). Image segmentation for fruit detection and yield estimation in apple orchards. Journal of Field Robotics, 34(6), 1039–1060.

Chu, P., Li, Z., Lammers, K., Lu, R., & Liu, X. (2021). Deep learning-based apple detection using a suppression mask R-CNN. Pattern Recognition Letters, 147, 206–211. https://doi.org/10.1016/j.patrec.2021.04.022

Chu, P., Li, Z., Zhang, K., Chen, D., Lammers, K, & Lu, R. (2023). O2RNet: Occluder-occludee relational network for robust apple detection in clustered orchard environments. Smart Agricultural Technology, 5, 100284. https://doi.org/10.48550/arXiv.2303.04884.

Chu, P., Li, Z., Zhang, K., Lammers, K., & Lu, R. (2024). High-precision fruit localization using active laser-camera scanning: robust laser line extraction for 2D-3D transformation. Smart Agricultural Technology 7, 100391. https://doi.org/10.1016/j.atech.2023.1

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. Computers and Electronics in Agriculture, 177, 105687. https://doi.org/10.1016/j.compag.2020.105687

Gene-Mola, J., Gregorio, E., Guevara, J., Auat, F., Sanz-Cortiella, R., Escola, A., …, & Rosell-Polo, J.R. (2019). Fruit detection in an apple orchard using a mobile terrestrial laser scanner. Biosystems Engineering, 187, 171–184. https://doi.org/10.1016/j.biosystemseng.2019.08.017

Kang, H., & Chen, C. (2020). Fast implementation of real-time fruit detection in apple orchards using deep learning. Computers and Electronics in Agriculture, 168, 105108. https://doi.org/10.1016/j.compag.2019.105108

Lazaros, N., Sirakoulis, G.C., & Gasteratos, A. (2008). Review of stereo vision algorithms: from software to hardware. International Journal of Optomechatronics, 2(4), 435–462.

Raguram, R., Chum, O., Pollefeys, M., Matas, J., & Frahm, J.M. (2013). USAC: A universal framework for random sample consensus. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(8), 2022–2038. https://doi.org/10.1109/TPAMI.2012.257

Xiong, Y., Peng, C., Grimstad, L., From, P.J., & Isler, V. (2019). Development and field evaluation of a strawberry harvesting robot with a cable-driven gripper. Computers and Electronics in Agriculture, 157, 392–402. https://doi.org/10.1016/j.compag.2019.01.009

Zhang, K., Chu, P., Lammers, K., Li, Z., & Lu, R. (2024). Active laser-camera scanning for high-precision fruit localization in robotic harvesting: System design and calibration. Horticulturae, 10, 40. https://doi.org/10.3390/horticulturae10010040

Zhang, K., Lammers, K., Chu, P., Li, Z., & Lu, R. (2021). System design and control of an apple harvesting robot. Mechatronics, 79, 102644. https://doi.org/10.1016/j.mechatronics.2021.102644

Zhang, K., Lammers, K., Chu, P., Li, Z., & Lu, R. (2024). An automated apple harvesting robot – from system design to field evaluation. Journal of Field Robotics, 41(7), 2384–2400. https://doi.org/10.1002/rob.22268