Efficient scene reconstruction for automated fruit maturity estimation

Stefan Baar, Yosuke Kobayashi, Kazuhiko Sato, Satoshi Kondo, Shinya Watanabe

Graduate School of Engineering, Muroran Institute of Technology, Muroran, Hokkaido, Japan

web: https://github.com/StefanBaar
e-mail: sbaar@muroran-it.ac.jp

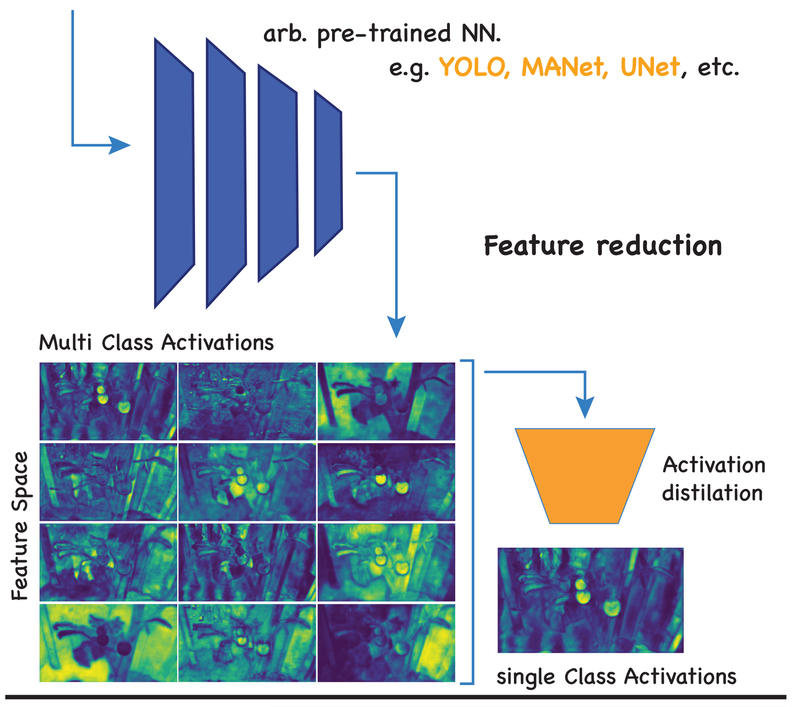
Abstract

We present a general deep-learning-based approach to estimating the degree of maturity of tomatoes from RGB images by estimating the individual tomato size from instance segmentation (IS) and depth estimation (DE). Our approach is based on the rotation symmetries exhibited by small tomato fruits (cherry tomatoes, Solanum lycopersicum var. cerasiforme), which permits the size estimation of the individual fruit even within a perturbed system. We show that the diameters and volumes of small round fruits can be evaluated with high precision (±5%) when using a set of known reference objects within the field of view. We evaluate our approach for single- and multi-frame RGB datasets. We were able to estimate the diameter from a single image, although with higher uncertainties than that from multi-frame estimation, due to the high uncertainties exhibited by the utilized instance segmentation and monocular depth estimation approaches and their down-stream error propagation.

Instance segmentation

Detect morphology of individual tomato fruit within an image by distilling the Multi Class Activations of a pre-trained Neural Network (NN) [3] to create a single class (tomato fruit) Activation function.

Data sample with detections

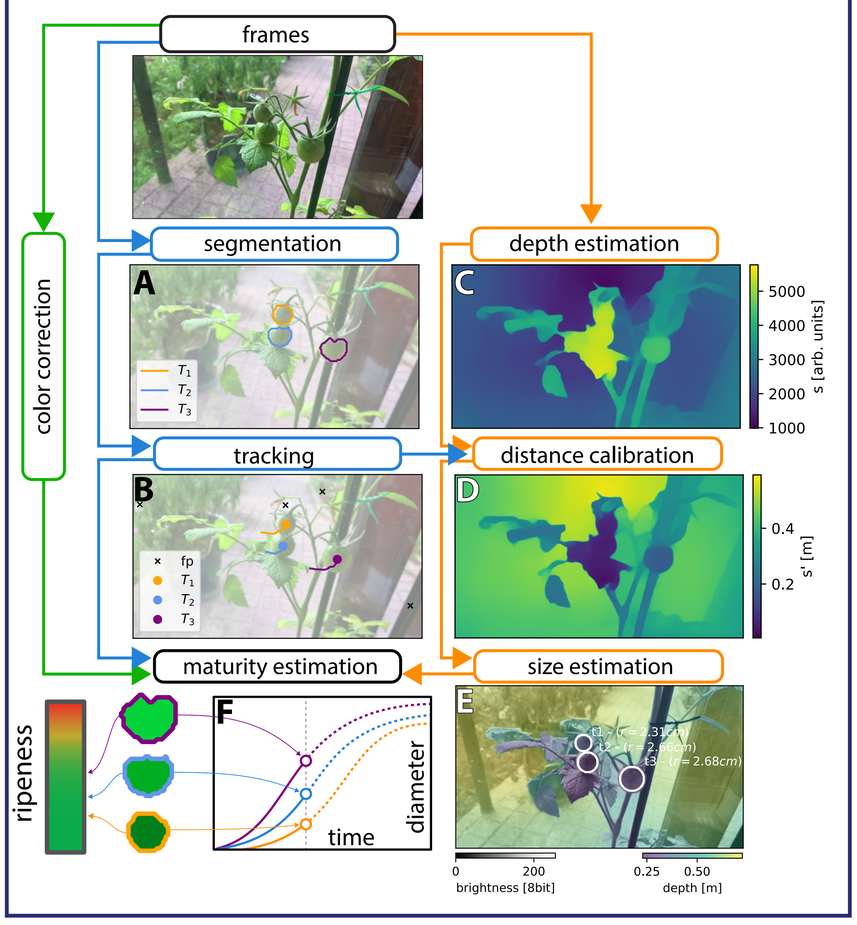


Introduction

- Aging society --> shortage of available labor for monitoring and management of large greenhouse environments [1]
- Automated fruit monitoring:
  - help reduce the need for physical labor
  - continuously monitoring and analyzing data on fruit development
  - informed decisions about irrigation, fertilization, etc.
  - improve efficiency, precision, and quality control, leading to better yields and higher-quality produce.
  - prediction of all above
- Regular fruit counting and size estimation can help estimate harvest yields and detect regularities early on by comparing the individual fruit characteristics to optimal growth conditions.
- Automatic Fruit size is estimation difficult [2]
  - requires: **3D scene reconstructin** and Instance Segmentation,

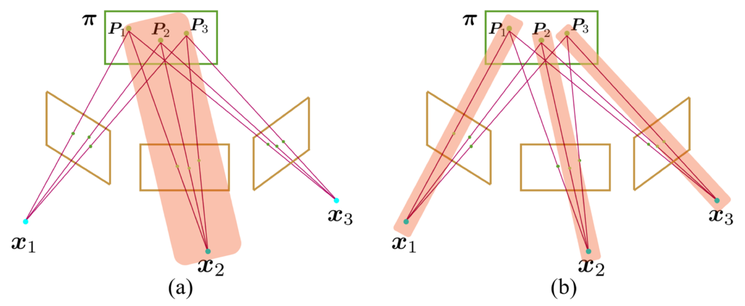
Methods

We combine instance segmentation and structure from motion to fully atomatically measure the fruit dimaeter from a sequence of RGB images. We compare OPTICAL FLOW and MONOCULAR DEPTH ESTIMATION methods for **STRUCTURE OF MOTION -> 3D IMAGE RECONSTRUCTION** on a set of optimized datapoints. The data point filtering (masking) is performed through NN based INSTANCE SEGMENTATION.



Structure from Motion

reconstruct a 3D scene and simultaneously obtain the camera poses of a monocular camera in a given scene.



- Feature Matching
  - MONOCULAR DEPTH ESTIMATION
    - Depth Anything
    - Midas
    - etc.
  - OPTICAL FLOW
    - Lucas Kanade
    - RAFT
    - etc.
- Outlier rejection
- Fundamental Matrix > Essential Matrix > Camera Poses
- Triangulation
- Bundle Adjustment
- Metric calibration (g - metric tensor)

From Fundamental Matrix F to Camera Pose

Solving following equations using Singular Value Decomposition (SVD).

$$\mathbf{x}_i'^{\mathrm{T}}\mathbf{F}\mathbf{x}_i = 0 \qquad [x_i' \;\; y_i' \;\; 1]\begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix}\begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = 0$$

$$x_i x_i' f_{11} + x_i y_i' f_{21} + x_i f_{31} + y_i x_i' f_{12} + y_i y_i' f_{22} + y_i f_{32} + x_i' f_{13} + y_i' f_{23} + f_{33} = 0$$

for a set of m points (N>8):

$$\begin{bmatrix} x_1x_1' & x_1y_1' & x_1 & y_1x_1' & y_1y_1' & y_1 & x_1' & y_1' & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_mx_m' & x_my_m' & x_m & y_mx_m' & y_my_m' & y_m & x_m' & y_m' & 1 \end{bmatrix}\begin{bmatrix} f_{11} \\ f_{21} \\ f_{31} \\ f_{12} \\ f_{22} \\ f_{32} \\ f_{13} \\ f_{23} \\ f_{33} \end{bmatrix} = 0$$

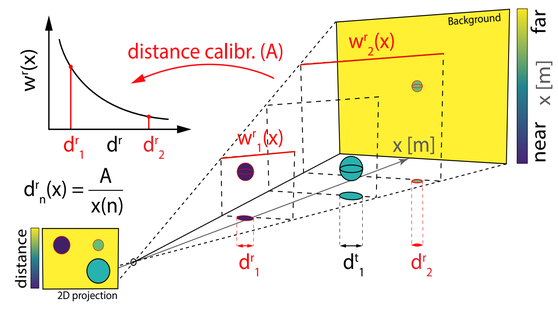
Non-Linear Triangulation

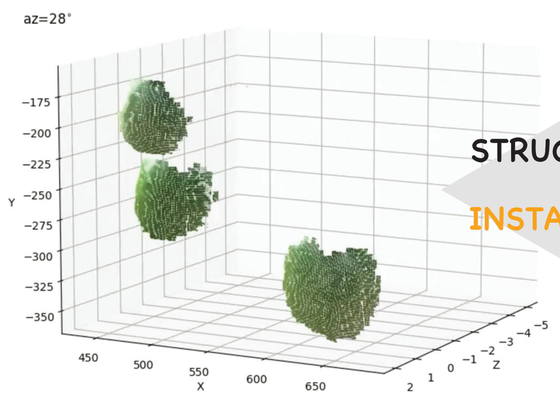
Given a set of camera poses (P) and linearly triangulated points (X), the locations of the 3D points that minimize the reprojection error can be refined. The linear triangulation minimizes the algebraic error. Though the reprojection error is a geometrically meaningful error, it can be computed by measuring the error between P and X as presented in equation 3, where, c is the index of each camera, X is the hoomogeneous representation of X. PnT is each row of camera projection matrix, P. The initial guess of the solution, X0, is estimated via the linear triangulation to minimize the cost function. [4]

$$\Delta_P^2 = \min_x \sum_{c=1,2} (u^c - \frac{P_1^{cT}\Phi}{P_3^{cT}X})^2 + (v^c - \frac{P_2^{cT}\Phi}{P_3^{cT}X})^2$$

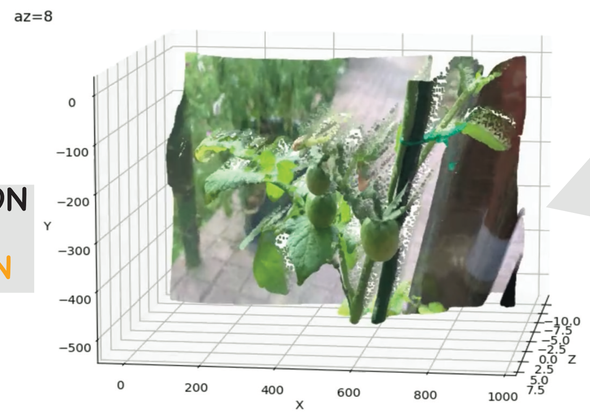
Metric claibration:

Metric of the scene is inherintly undefined (unit is pixel).Performed by comparing the scene to a known object (white sphere of known size) [5].





STRUCTURE FROM MOTION
+
INSTANCE SEGMENTATION



- Nonlinear but steady for OPTICAL FLOW
- but highly unsteady for MONOCULAR DEPTH ESTIMATION
- SIZE ESTIMATION for SINGLE IMAGES is very unstable und unreliable

Conclusion

- Fruit size estimation from RGB images is an unexpectatly complex and non-linear task.
- NN based segmentation and feature traking is still unreliable.
- **But statistical predictions are possible**

References

[1] U. Verma, F. Rossant, I. Bloch, Segmentation and size estimation of tomatoes from sequences of paired images, EURASIP Journal on Image and Video Processing 2015 (1) (2015) 1–23.

[2] M. Afonso, H. Fonteijn, F. S. Fiorentin, D. Lensink, M. Mooij, N. Faber, G. Polder, R. Wehrens, Tomato fruit detection and counting in greenhouses using deep learning, Fron- tiers in plant science 11 (2020) 571299.

[3] P. L. T. Mbouembe, G. Liu, J. Sikati, S. C. Kim, J. H. Kim, An efficient tomato-detection method based on improved yolov4-tiny model in complex environment, Frontiers in Plant Science 14 (2023) 1150958.

[4] J. Gené Mola, R. Sanz Cortiella, J. R. Rosell Polo, A. Escolà i Agustí, E. Gregorio López, Apple size estimation using photogrammetry-derived 3d point clouds, in: 9th Annual Catalan Meeting on Computer Vision. September 19, 2022, Universitat Autònoma de Barcelona,(http://acmcv. cat/), 2022.

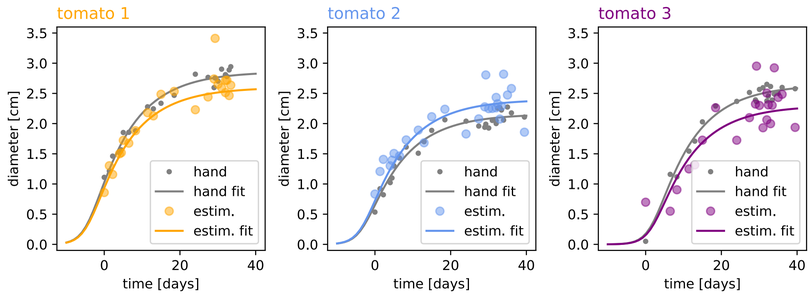
[5] J. C. Miranda, J. Gené-Mola, M. Zude-Sasse, N. Tsoulias, A. Escola, J. Arnó, J. R. Rosell- Polo, R. Sanz-Cortiella, J. A. Martínez-Casasnovas, E. Gregorio, Fruit sizing using ai: a review of methods and challenges, Postharvest Biology and Technology 206 (2023) 112587.

Prediction

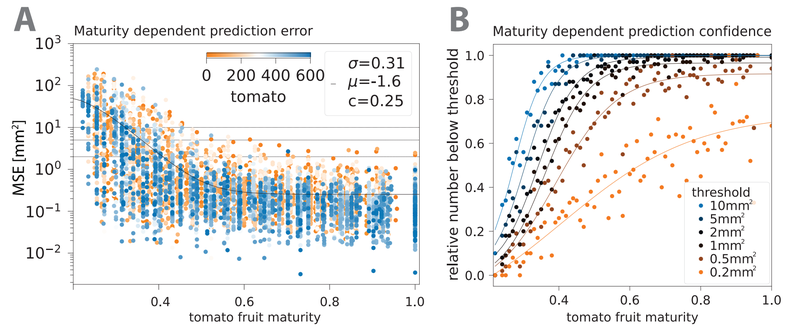
The Richards function is a more general (asymmetric) extension of the sigmoid function and was defined by Francis John Richards in 1959 with the following form shown in eq. 2, where A and a characterize the right horizontal asymptote. d will take the role of the growth rate for A=1, c=1,z=1 and b=1. For the application in the study, we assume C = 1 and either (A=1,a=0) or (A=const.,a=1), for which A and a-->A becomes the carrying capacity.

$$f_R(x) = a + \frac{A-a}{(c+be^{-d(x-x_0)})^{1/z}} \quad with \quad z > 0$$

The growth pattern of tomatoes one to three:



Tomato furit growth can be characterized by the Richards function. We have optimized the logistic function and reduced its paramters based on a dataset containing more 600 fruits. As presented below, we are able to measure the frut diameter with a precision < 3mm.



We defined maturity as a linear scale between the harvest time (maturity=1) and the time at which at least four data points were accumulated (maturity = 0.2). Low maturity was associated with fewer data points and higher measurement errors. The MSE exhibits a lognormal distribution, as shown below:

$$MSE(m) = I\frac{1}{m\sigma\sqrt{\pi}}e^{-\frac{(\ln m-\mu)^2}{2\sigma^2}} + c, \quad (0.2 \le m \le 1)$$