Automated Image Based Tomato Ripeness Categorization Using Python and the OpenCV Library

Alexander K Sweet[1,2], Manoj Sapkota[2], Lei Zhang[2], and Esther van der Knaap[2]

[1]Clarke Central High School, 350 S Milledge Ave, Athens, GA 30605
[2]Institute of Plant Breeding, Genetics, and Genomics at University of Georgia, 111 Riverbend Rd, Athens, GA, 30602

Introduction

Analyzing fruit phenotypic expression is critical to drawing conclusions from genetic experiments. Many computer-based systems have been used to improve the efficiency of fruit morphometric measurement; however, similar systems for ripeness are not as well developed. Neural networks have been used to categorize fruit ripeness, but these systems are complicated and time consuming to train and use. We propose a scalable pipeline to automatically, efficiently, and reliably categorize fruit ripeness by hue, saturation, and value (HSV). Using Python and the OpenCV and Pandas Libraries, the algorithm identifies tomatoes in images and averages their pixel's HSV values to yield a measure of each tomato's ripeness. This simple and automatic pipeline corresponds to current standard scales of ripeness (https://www.ams.usda.gov/grades-standards/tomato-grades-and-standards).

Results

While harvesting tomatoes for our experiments, we assumed that the ripeness of a mature tomato had no effect on its morphometry. From a total of 26 images, 183 harvested tomatoes were measured for fruit area and pericarp area using Tomato Analyzer software and visually classified for ripeness by color. Two one-way ANOVAs were preformed to compare the effects of color on tomato area and pericarp area measurements. The results showed that the color of a mature tomato has a significant effect on its fruit area, with largest the effect observed in the pericarp (Figure 1).

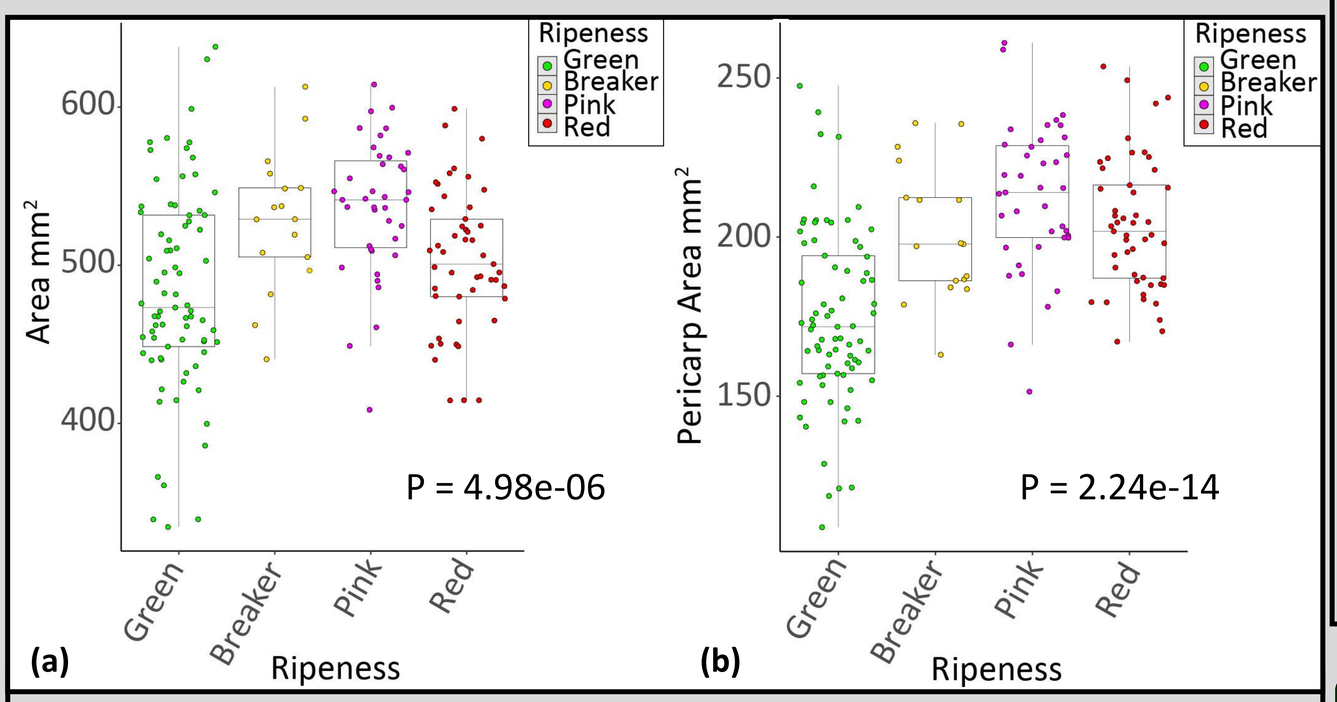


Figure 1. Total area (a) and pericarp area (b) of tomatoes in four color categories.

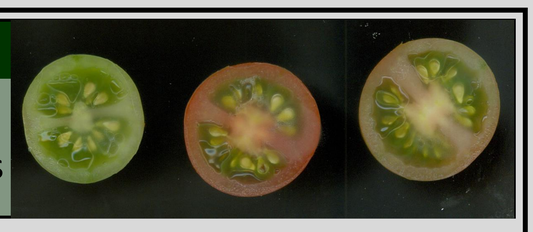
Due to significant effects, it was concluded that tomato ripeness as assessed by its color should be controlled in morphometry analyses.

Demonstration of the Novel Ripeness Assessment Method

A Python script using OpenCV and Pandas libraries was used to detect the same 183 tomatoes used in the preliminary experiment in 26 images and automatically categorize them by color to determine its ripeness (Figure 2).

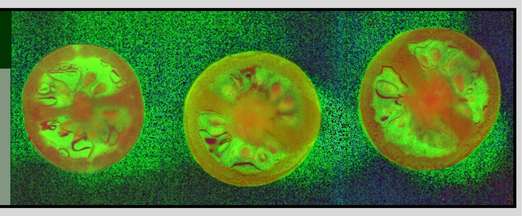
Document Scanner

Tomatoes are bisected and scanned to produce RGB images of their cross sections



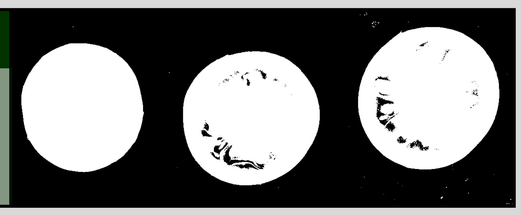
OpenCV

The original RGB image is converted into an HSV image



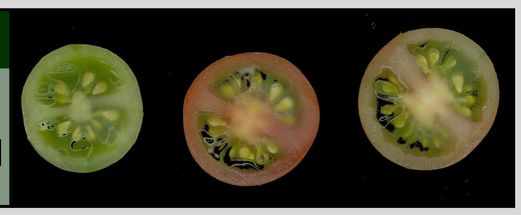
OpenCV

A HSV threshold is used to create a binary mask of the image



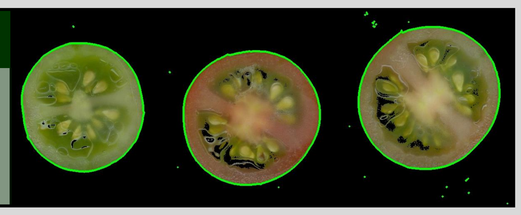
OpenCV

This mask is applied to the original image to separate tomatoes from the background



OpenCV

OpenCV's contour detection algorithm selects each tomato's shape in the image



OpenCV and Pandas

The HSV pixel values in each tomato's shape are averaged for each tomato and thresholded by Pandas as ripe or unripe

	Area	Image Coordinates		HSV			Output
	in pixels	X	Y	Hue	Saturation	Value	Category
Tomato 1	71007	2415	83	74	81	56	Unripe
Tomato 2	82809	2803	102	46	78	57	Ripe
Tomato 3	90279	3223	44	60	72	56	Unripe

Figure 2. Automatic ripeness determination pipeline and tabular output

The original picture of multiple tomatoes is processed to output a table containing each tomato's area, location in the image, average HSV, and classified ripeness

Conclusions, Implications, and Future Directions

Conclusions

- Variation in ripeness needs to be controlled when examining relationships between genetics and morphometrics.
- This novel method can be used to automatically, efficiently, and reliably determine ripeness.

Implications

- This open-source method can be used to complement Tomato Analyzer and other software to better capture genetic effects on tomato morphometric phenotypes.

Future directions

- Integrate this programming into Tomato Analyzer program
- Identify HSV thresholds to categorize tomatoes that do not ripen red.
- Develop portable ripeness determination systems for use in the field.

Acknowledgments

This research was funded by:

NSF IOS 154366 and NSF IOS 2048425