Measuring the size of strawberries using binocular photos

Kris van Melis
Supervisor(s): Junhan Wen, Thomas Abeel
EEMCS, Delft University of Technology, The Netherlands
20-6-2022

A Dissertation Submitted to EEMCS faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering

Abstract

To reduce food waste, it is important to know what strawberries to prioritise for harvesting. Size is an important quality attribute for strawberry. In order to know the size, the depth of the strawberry in the image must been known. To estimate the depth, stereovision gets utilized using binocular images. Since classic stereovision methods are quite inaccurate in predicting small areas in images, a technique by [Mustafah et al., 2013] is used. Using given segments of the strawberries, the left and right strawberry images will get matched. With the matched strawberries, the disparity can be calculated and thus the depth. Using the depth, the size can be estimated.

1 Introduction

The shape and size of strawberries are important factors that influence the quality of strawberries [Vittori et al., 2018]. For harvesting, it is important to know when strawberries are of optimal quality. Harvesting strawberries at the wrong time will increase food waste. We can reduce food waste by predicting the ripeness of the strawberry.

Currently, most current research uses color values to predict strawberry ripeness. [Brouwer et al., 2019] However, shape and size are also important factors that can help indicate quality.

To accurately predict the size of strawberries using photos, the depth must be known. Stereo vision can give depth information based on binocular images [Campbell et al., 2018]. With this depth information, the size of objects can be predicted more accurately [Mustafah et al., 2013].

No research has yet done depth and size estimation in non laboratory environments using binocular images. The aim of this work is to test whether size estimation of strawberries using binocular images is possible and how accurate it is.

To answer this question, the following sub-questions need to be answered: How can the depth of the strawberry be estimated? How can we measure the size of the strawberries themselves?

First, the methodology will be described. Then, an exploration of the available data will be done. After this, the pre-processing of the data is explained. The depth and size estimation is covered next, followed by the experiment setup and results. Finally, the results are discussed, a conclusion is given, and recommendations for future work are given.

2 Methodology

To measure an accurate size of the strawberries using photos, the depth of the strawberries in the picture must be known.

Depth can be acquired from the binocular view created by the camera and the near-infrared camera. Using the binocular view, the depth can be estimated using Stereo Vision.

Passive stereo vision works by calculating the disparity by matching the same points of the two images. However, to achieve good accuracy, the textures must be well defined and

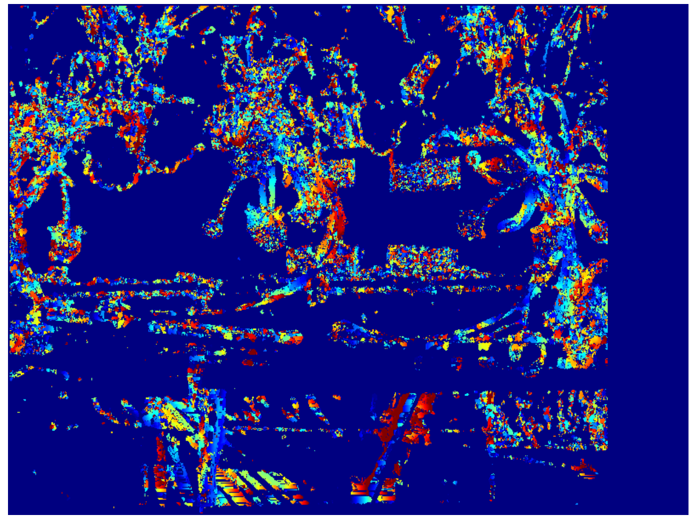

Figure 1: Classic Stereovision depth estimation on the binocular images given. Red indicates a lower depth, blue indicates a higher depth

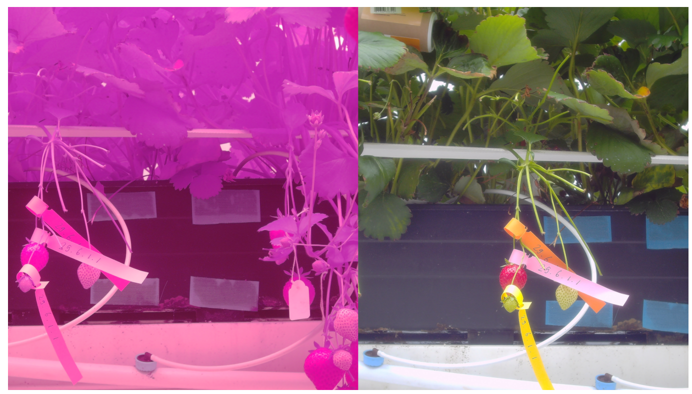

Figure 2: An example of the stereo image created by left the OCN camera and right the RGB camera

unique. Since most pictures will not have this, the results will be quite noisy. See figure 1.

To counteract this, a technique proposed by [Mustafah et al., 2013] will be used. Only the depth of the strawberries needs to be known. If the locations of the strawberries are known, only those locations need to be matched together from the left and right camera to get the disparity and thus the depth of the strawberries [Mustafah et al., 2013].

3 Data exploration

Given research by Junhan Wen and Thomas Abeel, stereo images with detected strawberries are available. The stereo image is formed by an RGB and an OCN camera, which are 10 centimeters apart. The use of an OCN camera is purely logistical. This has no further benefit to the research itself. See the example of the stereo images in figure 2. For both these images, the strawberries have been detected and their locations are known. See figure 3. In total there are 55 strawberries of which the size label is known and the corresponding left and

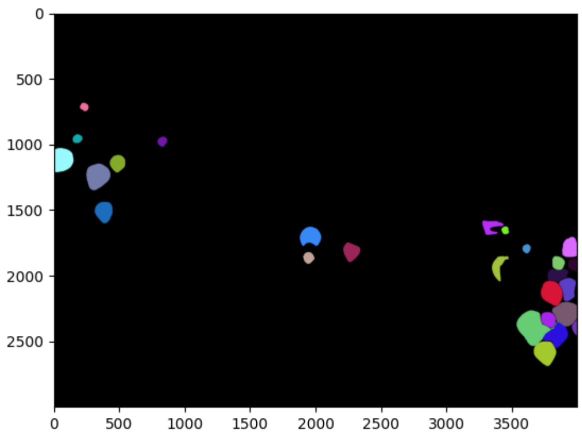

Figure 3: An example of the detected segments given

right images have been found. However, more strawberries are available without a size label.

4 Pre-processing of camera and segment data

Before the stereovision algorithm can be executed, the images must first be pre-processed. This is done using the following pipeline, also shown in Figure 4.

Camera Calibration The lens distortion is corrected by calibrating the camera. This is usually done by using a direct linear transformation [Xu et al., 2011]. This process has already been done for both the RGB and near-infrared cameras at setup, so the distortion is minimal.

Correction Transform After the camera calibration, an error in height must still be corrected by a transform. Both cameras must be aligned on the same height in order to get good stereovision results. Thus, a transform is used to correct for this.

Bounds analysis The bounds of the segments is needed to constrain the stereovision algorithm based on real-life constrains. In this step, based on the polygon information of the segment, the minimal and maximum bounds of the segment in the images gets calculated.

Polygon Area The area of the segment will be calculated on the basis of the given polygon. This will be used as a score to determine how many segments are alike.

UUID Assignment In order to keep track of all segments, a unique id gets assigned to every segment. This way, all segments in a pair of images have a unique id for tracking purposes.

5 Depth and size estimation

To measure the size of the strawberry, the depth must be known in order to obtain accurate measurements. Therefore, the depth estimation will be done first. After this, the size can be estimated. An overview of the algorithm as a whole can be found in figure 5

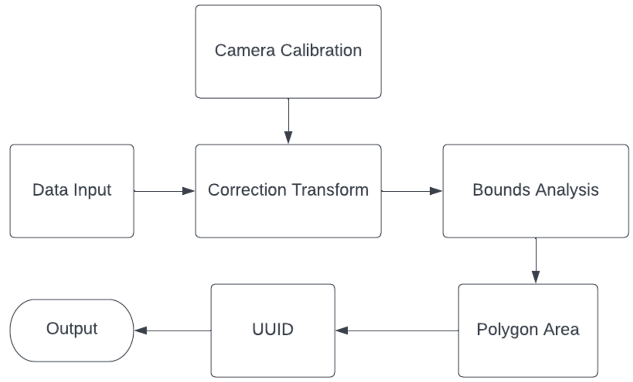

Figure 4: The pre-processing pipeline

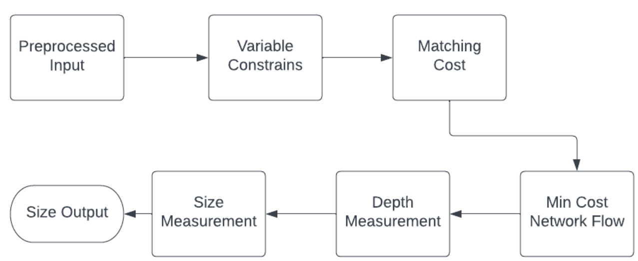

Figure 5: An overview of the depth and size estimation

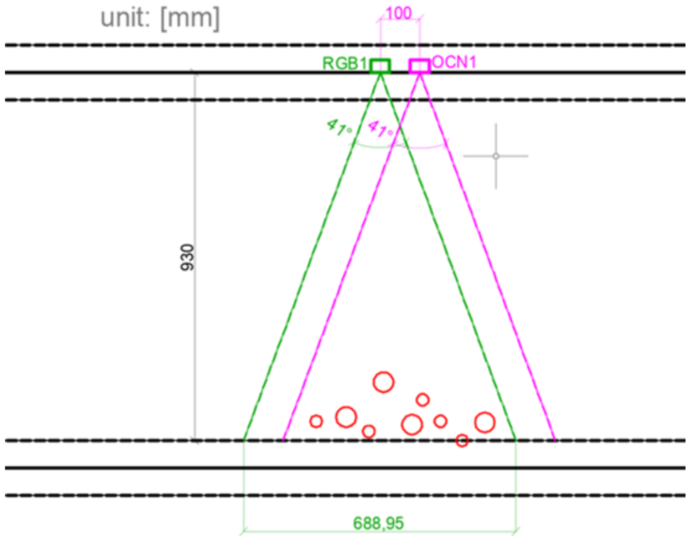

Figure 6: A blue print of the green house setup

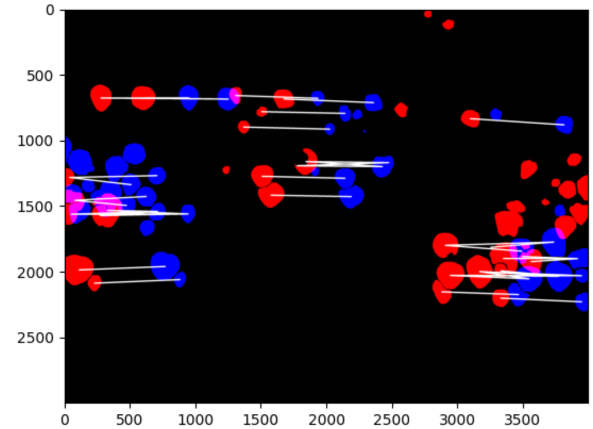

Figure 7: Potential Matches. Red segments are strawberries from the right camera. Blue segments are strawberries from the left camera. A white line indicates a potential match after variable constrains

5.1 Depth Estimation

After the pre-processing, the depth estimation will start. Here, the goal is to measure the depth of all segments. To do this, the segments of the left image must first be paired to those of the right image. This is done in 3 steps: variable constrains, cost analysis and Min-Cost Network Flow matching. After this step, the disparity between the left and right segments can be calculated to measure the depth of the strawberries.

Variable Constrains

Based on the green house setup given, segment matching can be constrained by real-life constraints. First of all, all segments that can be possible matches must be on the same epipolar line. So only segments on the same horizontal line can be a potential match. Further, in the setup, it is shown that the maximum depth is no more than 103 cm and the minimum depth is 73 cm. Since there is 10 cm between both cameras, the maximum and minimum disparities can be calculated. Using this information, potential matches can only be on the same horizontal line and must be offset between the minimum and maximum disparity. These values can be calculated using the green house setup shown in Figure 6. A result of the variable constrains can be found in figure 7.

Matching Cost Function

After constraining potential matches, more similar segments make better matches. To prefer matches that are more similar, a cost function is chosen. It is important for this cost function to be fast, as many potential matches must be compared. The polygon area is a good function since it is quick and gives meaningful data on how much the segments are similar. The cost is calculated as the percentage of difference between the segments.

$$cost = abs(area_l - area_r) / \\ (area_l + area_r) * 100 \tag{1}$$

Min-Cost Network Flow

To match the segments as best as possible, the Min-Cost Network Flow is used to make as many matches as possible with the least total cost. Segment matching is a maximum bipartite problem. A segment can only be matched with one other segment of the other camera and is restricted by potential matches.

However, some matches might be better than others. That is why the minimum cost is also important, since this will give us the most and best matches possible. As a cost function, the polygon area is used.

This problem can be modeled in the Min-Cost Network Flow in the following way:

First, a source and sink node is created. For the source and sink, supply and demand will be set to the maximum possible matches that can be acquired. The maximum possible matches that can be acquired is the minimum number of segments of both cameras.

After this, a node is made for all segments from the left camera and connected to the source. The capacity of these connections is 1, since a segment can only be matched once. The same process is done for all segments from the right camera, however these will be connected to the sink.

Now, for all potential matches between the left and right, an edge is made with a capacity of 1, since a segment can only be matched once. On this edge, a cost is added. This cost is the count of pixels as discussed in Matching Cost Function 5.1. However, if the cost is higher than 70%, the edge is removed, as matches that are very different are not desirable. For an overview, see Figure 8

Finally, the Min-Cost Network Flow is solved and returns the most number of matches with the minimum amount of cost.

Depth Measurement

Now that the segments have been matched, the depth can be measured by the disparity between the left and right seg-

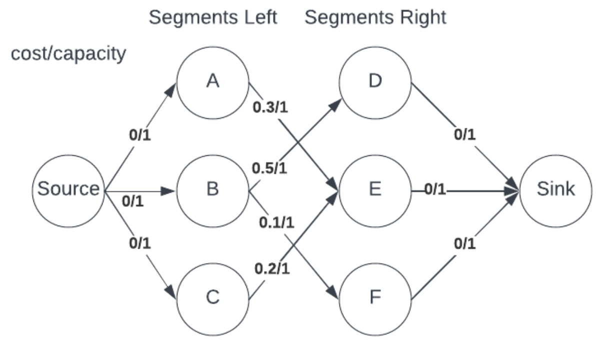

Figure 8: An example network model for matching the left and right segments

ments. The disparity is calculated by taking the center of both segments and calculating the disparity in pixels. Using the field of view and the image width in pixels of the camera, the depth is calculated using the following formulas:

$$focalPixel = (imageWidth * 0.5) / \\ tan(FOV * 0.5 * \pi/180) \quad (2)$$

$$depth = baseline * focalPixel/disparity \quad (3)$$

Where in the current setup, the FOV is 41 degrees, the image width in pixels is 4000 pixels and the baseline is 10 cm.

5.2 Size Estimation

After knowing the depth of each strawberry, we can estimate its size. Sizing measurements are done using three categories: tiny (< 25mm) , small ($25 - 30$ mm) and coarse (> 30 mm). So, the exact size itself is not important. The strawberries must only be categorized correctly in those categories. With a known depth, the real world distance covered by one pixel can be calculated. This is done using geometric similarity and a calibration point. Based on the setup, it is known that at 93 cm, the 4000 pixels cover 68.895 cm in real life. Using these measurements, the distance per pixel can be calculated for any depth.

$$constant = depth/length \quad (4)$$

$$lengthPerPixel = (depth/constant)/imageWidth \quad (5)$$

Now that the length per pixel is known, the number of pixels has to be measured. Since strawberries sometimes are occluded from view, the longest width from both segments is used. The hypothesis is that this measurement is accurate enough to assign the right category to the strawberry.

6 Experimental Setup and Results

To measure whether the method works, some experiments will be performed. Sadly, the only verification data that is available is the sizing categories itself. Thus, the depth estimation cannot be checked against absolute values. Instead, the matching algorithm and the relative depth estimation will be tested.

The algorithm gets tested on data from a real strawberry farm. Here, the camera setup has been created and over the course of a few months data have been collected. For 55 strawberries, we have the true size label data and both left and right segments.

6.1 Segment Matching Accuracy

In order to get good depth estimation, the matching of the left and right segments must be accurate. To test this, a human will annotate which segments from the left and right camera match with each other. Then, the algorithm will predict which segments belong together. The accuracy will be determined by the percentage of correctly matched segments. If a segment is not matched or is incorrectly matched, this is counted as an incorrect match.

Of the 172 pairs, 150 pairs have been matched correctly. This means that the pair matching algorithm has an accuracy of 87%.

6.2 Relative Depth Accuracy

Since there is no absolute data on the depth is available, the depth estimation will be tested using a ground truth of the relative depth created by humans. This ground truth will tell which strawberries are perceived to be closer than others. Then the prediction of the depth estimate will be checked against this. Every strawberry that has an incorrect relative depth will be marked as an error. An incorrect relative depth means that, the strawberry is wrongfully marked as in front or behind another strawberry. An accuracy score will be given on the basis of the percentage of correct matches.

Of the 173 segments reviewed of 11 images, 52 errors were found in relative depth. This means the relative depth estimation is 70% accurate.

6.3 Size Prediction Accuracy

Finally, the overall accuracy of the algorithm will be measured using the ground truth size labels. For 55 strawberries, the size category is known. The algorithm will be run on the segments and test whether it will give the correct size category. On the basis of the number of correct category labels given, an accuracy score is calculated as a percentage.

To see whether the depth estimation has a positive effect on the results, the algorithm will also be run with a constant depth of 900mm (an average depth for the strawberries) to see whether the depth estimation improves the accuracy.

Running the result on 55 strawberries that have been given a size category, 40 will be size labeled by the algorithm. Some strawberries cannot be labeled because there are no matching left and right segments for depth prediction. It could be that this segment is missing or not visible on both cameras. Some segments have been missing on one of the two cameras causing the size estimation to fail. Other segments were not visible from either camera.

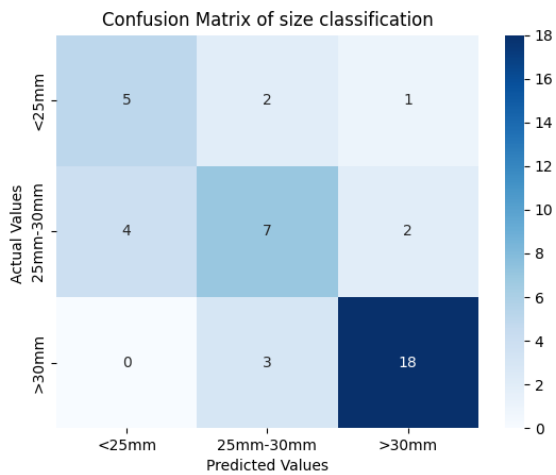

Figure 9: A confusion matrix showing the predicted and actual values of the size labeling with a constant depth of 900mm

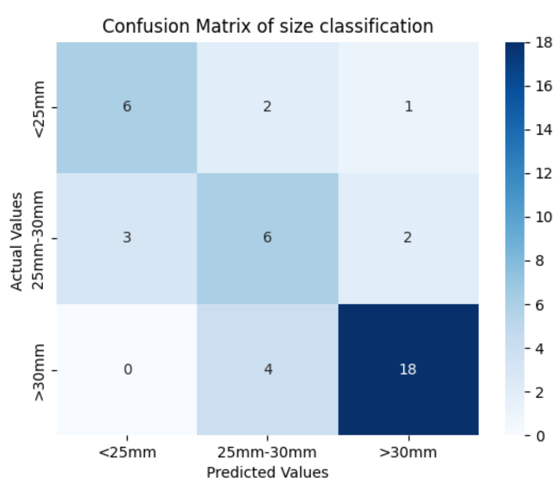

Figure 10: A confusion matrix showing the predicted and actual values of the size labeling using depth estimation

The baseline result with constant depth has matched 30 of the 40 strawberries correctly. A confusion matrix of these results can be found in figure 9.

Of the 40 strawberries that have been size-labeled, 30 have received the correct label. A confusion matrix of these results can be found in Figure 10.

7 Responsible Research

Based on the research topic, no ethical aspects are to be discussed. The research does not influence ethical values. The research can easily be reproduced using the code and images data from the TU Delft. However, as of now, it is not known whether all data can be published.

8 Discussion

The results might not be really accurate, as only a few data points were available to properly test the algorithm. This could be improved by collecting more data to test on, such as depth.

Furthermore, the data only consisted of one set of cameras in a green house in The Netherlands. Differences in results might occur when using other cameras or green houses.

Finally, the matching of pairs works well if all segments are given correctly. Missing or wrong segments could decrease the overall performance of this algorithm.

9 Conclusions

Based on the results, measuring the size of strawberries using binocular photos is possible with an accuracy of 70% percent if given a size label. 27% of the strawberries have not received a size label due to incomplete left and right segments. This can be the result of strawberries not being visible on both cameras, not being recognized, or segments being incorrectly matched.

However, comparing the size labeling to the base line, it seems the depth estimation has no real effect in improving the results.

The size prediction based on the three categories has a possible side effect that the absolute size prediction could be quite inaccurate. However, in the results, in general it is accurate enough to distinguish the three size categories.

10 Future work

From the conclusion, it is clear that more research and testing is needed to provide more accurate data on algorithm performance. Especially depth data and more sizing data is needed to further test and increase the performance of the algorithm. The depth data could function as the ground truth for the depth estimation algorithm. In this way, the depth estimation can be tested separately.

More sizing data and more accurate sizing data could fully test the algorithm as a whole. At the moment, only size labels are available. More meaningful results can be acquired by having absolute measures in millimeters. In this way, the real accuracy of the size estimation can be given.

References

[Brouwer et al., 2019] Brouwer, B., Mensink, M., Woltering, E., and Silva, F. P. D. (2019). Predicting strawberry shelf life based on input quality.

[Campbell et al., 2018] Campbell, S., Krpalkova, L., Riordan, D., Walsh, J., Murphy, A., and Ryan, C. (2018). Computer vision for 3d perception a review.

[Mustafah et al., 2013] Mustafah, Y. M., Azman, A. W., and Ani, M. H. (2013). Object distance and size measurement using stereo vision system. volume 622, pages 1373–1377.

[Vittori et al., 2018] Vittori, L. D., Mazzoni, L., Battino, M., and Mezzetti, B. (2018). Pre-harvest factors influencing the quality of berries. *Scientia Horticulturae*, 233:310–322.

[Xu et al., 2011] Xu, G., Li, X., Su, J., Pan, H., and Tian, G. (2011). Precision evaluation of three-dimensional feature points measurement bybinocular vision. *Journal of the Optical Society of Korea*, 15:30–37.