Extracting the Dimensions of an object using a Stereo Vision System.

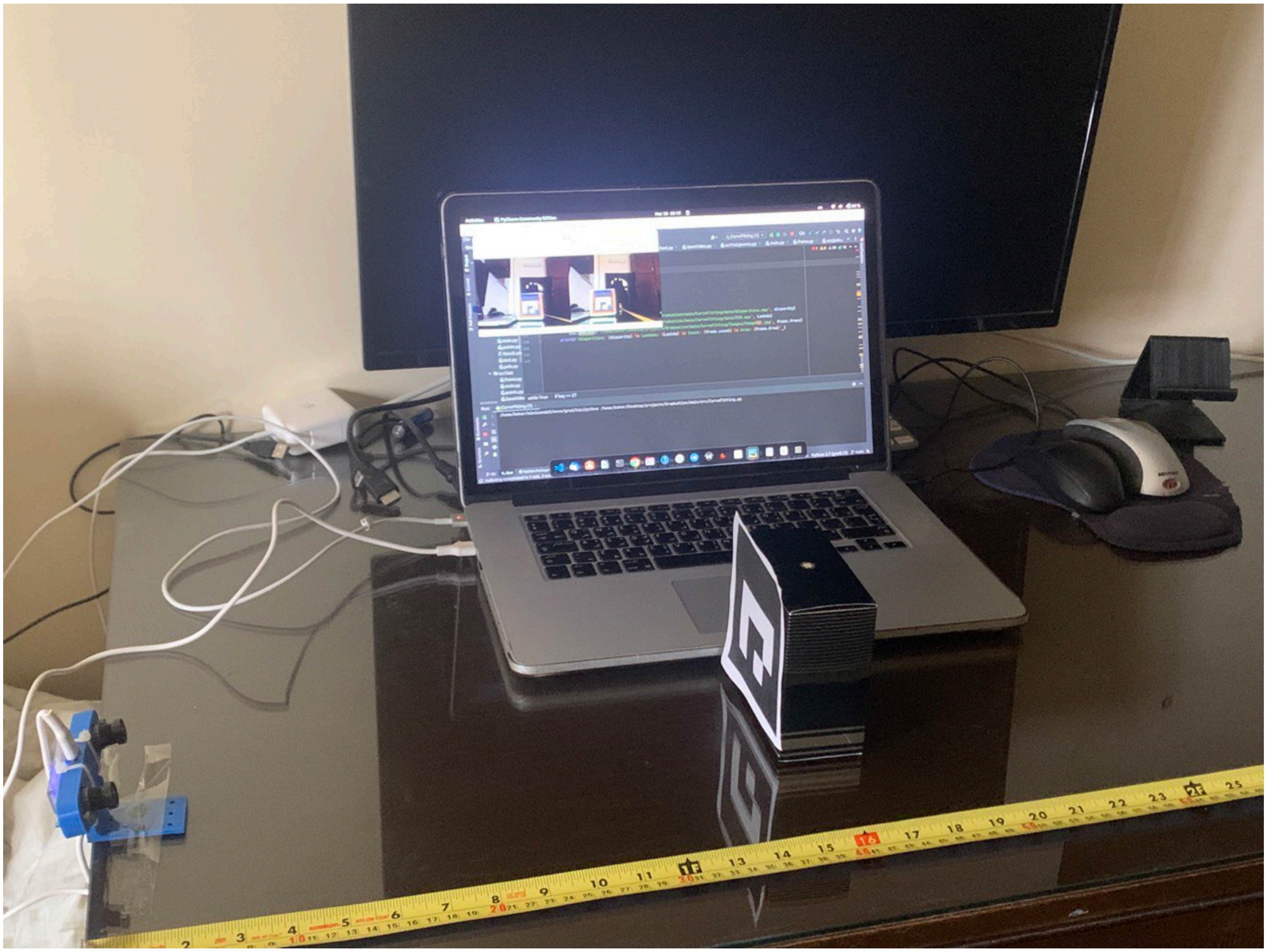
1. Introduction

The Stereo Vision Camera enables us with two different perspectives of the same object, therefore depth estimation is possible by computing the disparity of the object in both frames. For extracting Dimensions we don't focus on calculating the depth exactly, rather extract a relationship between the disparity and a measured length in the detected object. If we presume that (l is the length in pixels and L is the actual Length in cm) Then we can say that:

$L = \alpha l$ where α is a coefficient that represents the proportional factor between the measured length in pixels and the actual length. The goal now is to find a relationship between the disparity D and the coefficient α that we can utilize in realtime to convert lengths measured in pixels to cm.

2. Data Collection

We'll use an Aruco Marker which Is basically a distinct figure that is easy to detect and track in camera frames, similar to a barcode. The width of the Aruco marker is measured and used as the reference length in our experiment, this is the Length L. The stereo camera is mounted in a fixed spot and the marker is placed 5cm in front of the camera. Initially the Disparity (D) is Calculated from the two frames that contains the detected marker, the value of the disparity is noted down. After that, we will measure the width of the detected marker in one of the frames, this length is l. Now in this case, for the specific value D we can compute α using the equation above. Now we have our fist point (D,α). The distance of the marker from the camera is incremented by 1 cm each time and the same process is repeated. The accumulated points are then plotted on a graph and a using a Regression Algorithm we will extract the function that represents the relationship between our two independent variables, disparity and α coefficient.

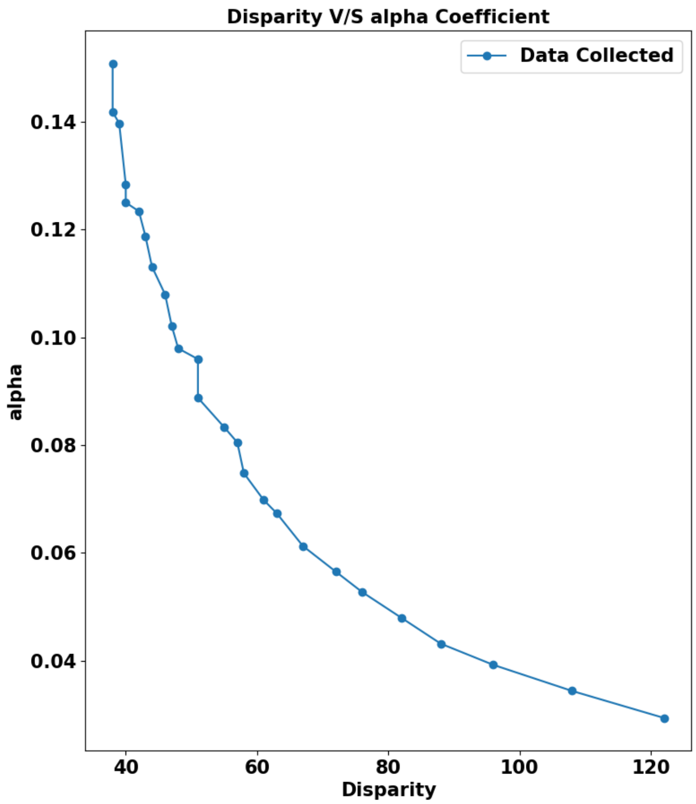


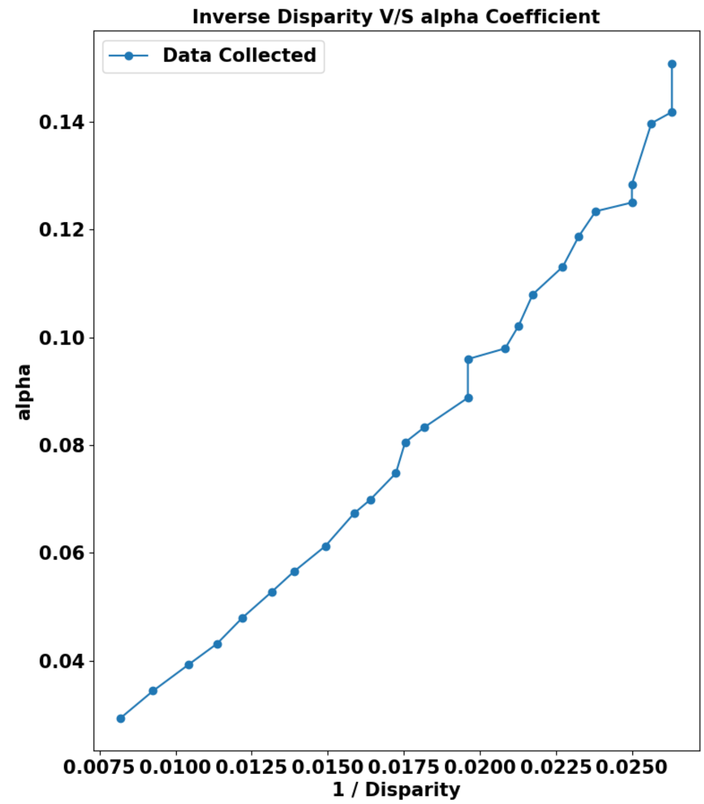
3. The Results

The following table represents the data collected.

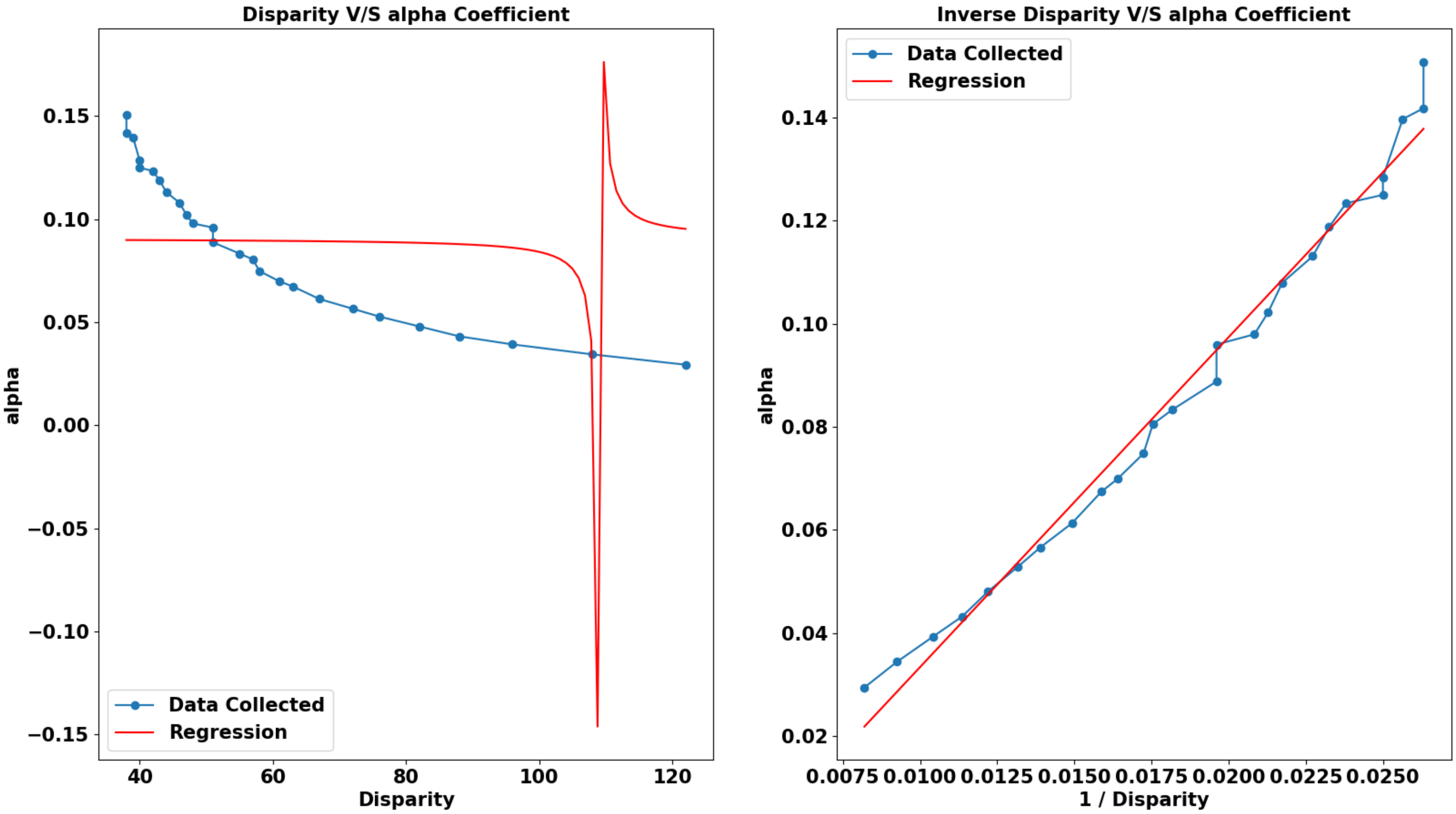
Depth [cm]	Disparity [pixel]	Alpha [cm/pixel]
7	122	0.0294
8	108	0.0344
9	96	0.0392
10	88	0.0431
11	82	0.04797
12	76	0.05277
13	72	0.0565
14	67	0.0612
15	63	0.0673
16	61	0.0698
17	58	0.0748
18	57	0.0805
19	55	0.08333
20	51	0.08878
21	51	0.0959
22	48	0.0979
23	47	0.1021
24	46	0.1079
25	44	0.1130
26	43	0.11875
27	42	0.1233
28	40	0.125
29	40	0.1284
30	39	0.1397
31	38	0.1417

Depth [cm]	Disparity [pixel]	Alpha [cm/pixel]
32	38	0.1507





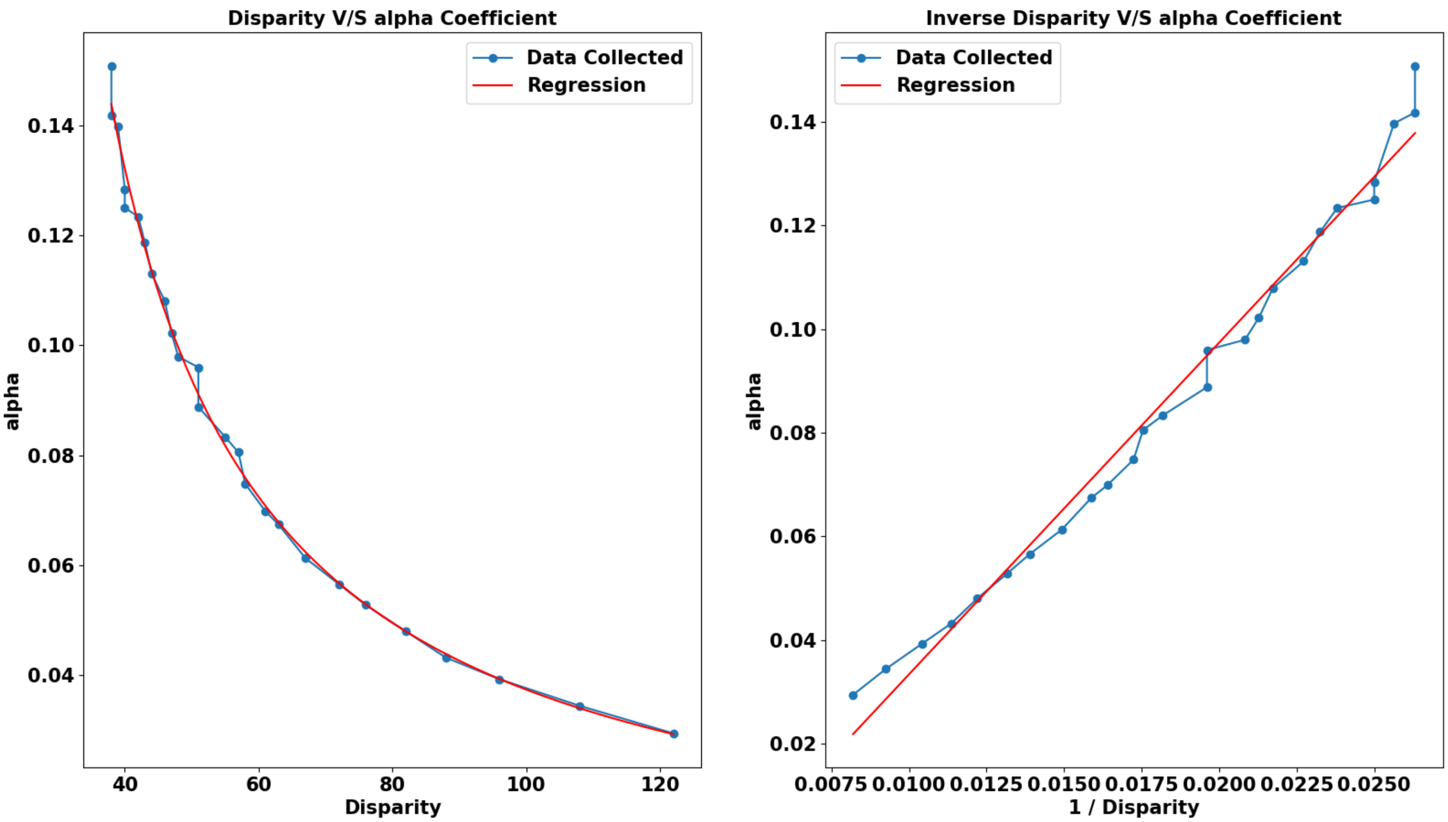
The above plot Visualizes the data collected, the next step is to Preform Regression on these points. To extract the function that represents the relationship between the two variables, we need to Curve Fit these points (Inverse Disparity V/S alpha is linear, So standard linear regression will do) . The conventional way of performing curve fitting is using the least squares method, this method is considered the most common. The least square method is a mathematical procedure for finding the best-fitting curve to a given set of points by minimizing the sum of the squares of the offsets ("the residuals") of the points from the curve. I won't go in-depth on this method because it won't work here, as we can see in the following plot:



Linear Regression works like a charm, but the Curve fitting is terrible. The Reason for that is the number of observations in our data is less than the number of variables, and that's because we want to fit our data to the function $f(x) = \frac{a}{x-b} + c$. So instead, we will have to use either the dogbox or the trf method for optimization. The dogbox algorithm does trust-region iterations, the shape of trust regions is rectangular as opposed to conventional elliptical. The intersection of a trust region and an initial feasible region is again some rectangle. Thus, on each iteration a bound-constrained quadratic optimization problem using a specific quadratic formula.

- <u>How the The Curve Fit function was chosen</u>: we can take reference on the relationship between disparity and depth. It's known that $depth = \frac{bf}{disparity}$ where b is the baseline that represents the horizontal distance between both lenses of the stereo vision camera, and f is the focal length. The relationship between the disparity and depth is a reciprocal function of $f(x) = \frac{a}{x}$, so for our data we try to fit the curve with a more generalized reciprocal function of $f(x) = \dfrac{a}{x-b} + c$ using the dogbox method mentioned above.

The Result of the curve fitting algorithm using the dogbox method and the function $f(x) = \dfrac{a}{x-b} + c$ as input, we get the following result:



It's clear that the results are much better and more reliable to use than the previous output.

The output function for the (Disparity - alpha) data is:

$$alpha = \frac{3.34713891}{disparity - 15.0538816} - 2.02645145e - 03.$$

Whereas the linear function for the (Inverse Disparity - alpha) data is

$alpha = 6.402 * (1/disparity) - 0.03065.$

4. Conclusion

The output function is reliable in computing lengths based on disparity and it will be used in our project for extracting size of fruits to use as one criteria to determine maturity.