Article

A Distance Measurement Approach for Large Fruit Picking with Single Camera

Jie Liu [1,2,*], Dianzhuo Zhou [1], Yifan Wang [1], Yan Li [1] and Weiqi Li [1]

[1] College of Engineering, Huazhong Agricultural University, Wuhan 430070, China;
zhoudz@webmail.hzau.edu.cn (D.Z.)

[2] Key Laboratory of Agricultural Equipment in Mid-Lower Yangtze River,
Ministry of Agriculture and Rural Affairs, Wuhan 430070, China

[*] Correspondence: liujie@mail.hzau.edu.cn; Tel.: +86-156-2326-3208

Abstract: Target ranging is the premise for manipulators to complete agronomic operations such as picking and field management; however, complex environmental backgrounds and changing crop shapes increase the difficulty of obtaining target distance information based on binocular vision or depth cameras. In this work, a method for ranging large-sized fruit based on monocular vision was proposed to provide a low-cost and low-computation alternative solution for the fruit thinning or picking robot. The regression relationships between the changes in the number of pixels occupied by the target area and the changes in the imaging distance were calculated based on the images of square-shaped checkerboards and circular-shaped checkerboards with 100 cm^2, 121 cm^2, 144 cm^2, 169 cm^2, 196 cm^2, 225 cm^2, 256 cm^2, 289 cm^2, and 324 cm^2 as the area, respectively. The 918 checkerboard images were collected by the camera within the range from 0.25 m to 1.5 m, with 0.025 m as the length of each moving step, and analyzed in MATLAB to establish the ranging models. A total of 2448 images of four oval watermelons, four pyriform pomelos, and four oblate pomelos, as the representatives of large fruit with different shapes, were used to evaluate and optimize the performance of the models. The images of the front were the input, while the imaging distances were the output. The results showed that the absolute error would be less than 0.06 m for both models and would linearly increase with a decrease in the distance. The relative error could be controlled at 5%. The results proved the proposed monocular method could be a solution for the ranging of large fruit targets.

Keywords: monocular vision; ranging; large-size fruit; checkerboard; model

1. Introduction

The picking robot is an effective way to deal with the current challenge of fruit harvest [1–3]. Machine vision plays the role of eyes for robots, aircraft technology, the Internet of things, and other mechatronics systems by providing the space coordinate figure [4]. With the location information of the target fruit, the picking robot can perform the supposed work [5,6]. As one of the machine vision methods for target distance assessment, monocular distance measurement has been applied in all sorts of fields, such as industrial robots [7], medical treatment [8], vehicle control [9], and fruit picking [10,11]. Compared to binocular vision and laser vision, it was more convenient and simpler and had great further development potential [12,13].

Initially, it was in the 1960s that research began on monocular vision. Lawrence Roberts at the Massachusetts Institute of Technology (MIT), as the ancestor of monocular vision, obtained a big hit due to the robot's intelligent recognition through simple building blocks [14]. The following decades witnessed a huge leap again in the monocular field, which paved the way for further development. In 1997, [15] achieved distance calculation according to the proportion of the object's projection on two images; one was taken at the camera point, while the other was taken at the point where the camera moves along the

optical axis. In 2005, [16] proposed a new ranging method in South Korea that used a single camera and a rotating mirror to obtain a series of reflected images through the camera in front of the rotating mirror and extract distance information from them.

Over the course of the last decade, with the boom of new technology, particularly the involvement of computers in machine vision, monocular distance measurement has stepped up to a higher level. In 2013, [17] designed an improved method that used a monocular camera to track specific objects and then employed an efficient computing algorithm implemented in MATLAB/SIMULINK to measure the distance from the object to the camera. In 2017, [18] gave a strategy—a target ranging method for a wheeled mobile robot based on monocular vision—that extended the ranging target from a plane object to a three-dimensional object and improved the measurement accuracy without correction.

In the last 5 years, [19] has used the monocular vision technology of instance segmentation and camera focal length to detect the absolute distance between the front camera and the car camera and has largely improved the measurement accuracy. At the same time, [20] thought the convolutional neural network could be applied to determine the depth from a single camera image; therefore, eight different networks were designed for depth estimation, each of which could suit a certain characteristic level. In 2019, [21] theoretically explored the relationship between local feature correspondence and the similarity of two images. The object will be identified based on a two-stage template matching process, and its proportion level will be determined at the same time. Then it will be successfully applied to the real-time automatic packaging sorting line. In the same year, [22] trained a fully convoluted neural network to carry out distance estimation, which predicted the size of different objects depicted in the image pixels. Although these methods have been discussed a lot about the monocular camera's distance measurement appliance in great fields, in the published research, they were rarely referred to for complex orchard environments and fruit surface imaging distortion.

In general, the technique of measuring distances with only a single camera has become more sophisticated and advanced and has achieved a historical transfer and breakthrough from the traditional simple geometric calculation to some new technologies such as computer image cognition and neural networks. As a fact, fruit trees always live in a complicated environment, and almost all of these technologies cannot be well adapted because of either excessive reliance on high-precision physical instruments or the high cost of time and energy. As a result, a simpler and more practical ranging method will be suitable for the orchard environment in view of the complex background in the images.

The application of monocular vision to image-based visual servoing (IBVS) required consideration of the impact of imaging errors, such as lens and optical paths, on detection accuracy and the problem of targets moving out of the camera's imaging area. Therefore, on the basis of camera calibration and internal and external parameters, this study proposed a model for distance prediction by the change in the number of target pixels in a continuous image acquired during the continuous motion of a single camera, which was used to obtain depth information of large fruit on a tree in a complex background.

The model used the change in the number of pixel values of the target area in the resulting image as an input variable, which allowed the inherent error of the system to be eliminated in the calculation process, and the process of continuous imaging with small values and fixed interval steps prevented the target from leaving the camera imaging area, thus providing a less computational and less costly technical solution for target distance prediction and operational action control of fruit picking manipulators. Although the application of monocular vision in fruit and vegetable picking was rare, its imaging system was simple, lightweight, small in size, did not require precise image matching, and would have advantages in fast computing speed, which not only enabled the cost of subsequent machine vision servo systems to be controlled within an acceptable price range but also facilitated the development of multi-robot collaborative working machines and could provide a reference for the identification of target depth information in the production process of other fruits and vegetables or crops.

In this work, the principle of monocular distance measurement was first studied and the materials and methods used were introduced in Section 2; then, in Section 3.1, the results of establishing models using square checkerboards and circular checkerboards of different areas were shown, while the fruit pictures of 2 oblate pomelos, 2 pyriform pomelos, and 2 oval watermelons were made to verify the accuracy of the 2 established models; Via error analysis, the two established models were modified to reduce the error, and the results are shown in Section 3.2. Finally, the whole work and the future expectations were concluded in Section 4. The flowchart of this work is shown in Figure 1.
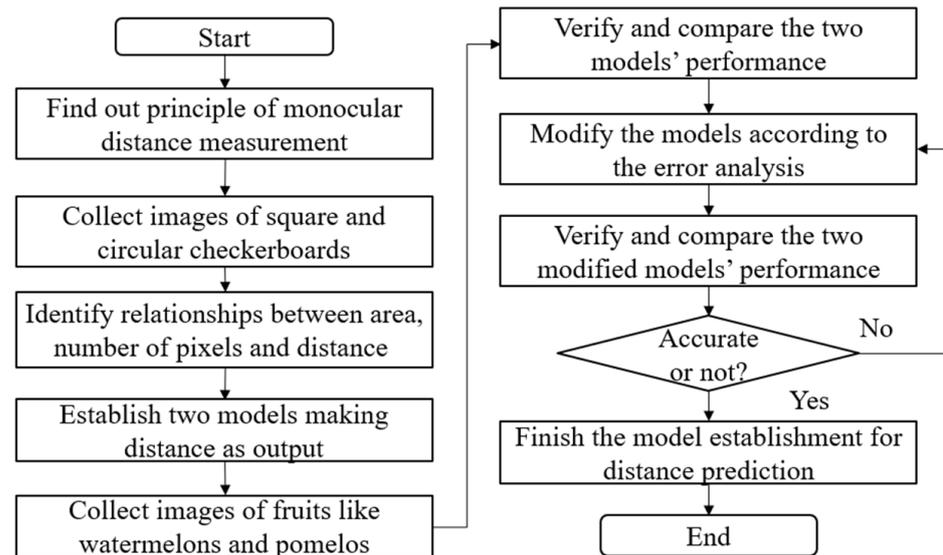

Figure 1. The flowchart of the main work.

2. Materials and Methods

2.1. Principle of Monocular Distance Measurement

The monocular distance measurement was mainly achieved depending on all captured information from images collected by the camera. The principle of camera shooting was presented in Figure 2. The reflected light from the photographed subject was concentrated by the camera lens to form a real image in an area between one and two times the focal length from the optical center. When the shooting distance *D* was far greater than the focal length *F*, approximately the image distance would be seen as equal to the focal length.

$$\frac{1}{F} = \frac{1}{D} + \frac{1}{V} \tag{1}$$

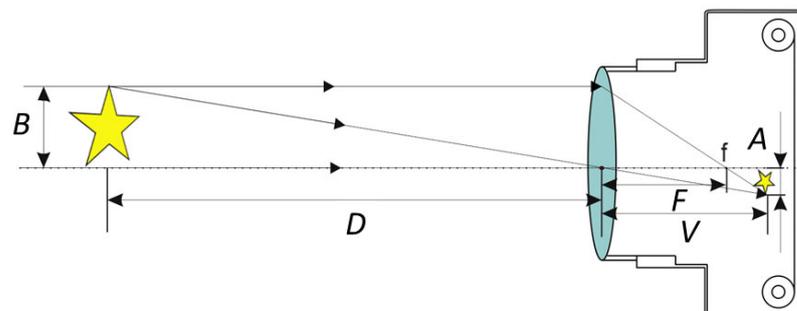Among those, *V*—image distance, *D*—the shooting distance, and *F*—the focal length.

Figure 2. Principle of camera shooting.

According to the lens imaging principle and the principle of a similar triangle, it could be:

$$\frac{A}{F} = \frac{B}{D}$$

(2)

A was the size of the real image formed in the camera, and B was the size of the shot object.

It is proposed that there are two patterns with the same area: a square of length L and a circle of radius R. As a result, the area S of the two patterns would be expressed as follows:

$$S = L^2 = \pi R^2$$

(3)

S represented the area, L represented the length of the square, and R represented the radius of the circle.

According to Equations (2) and (3), when a camera takes images of two patterns right from the front, the imaging areas would be expressed as:

$$s = \frac{F^2}{D^2} \cdot S$$

(4)

where s was the area of the real image formed in the camera.

On account of the shooting principle of the camera, the real image could be deemed an orderly array of pixels on the camera screen through the photoelectric converter and processor. Supposing the number of pixels of the video in the camera was x, the following equation described the special link between the imaging area of the objective and the number of pixels:

$$x = \omega s$$

(5)

where ω represents a constant decided by the camera, x was the number of pixels of the image in the photo.

Therefore, conversion from number of pixels to distance was performed with the following simplification of the formula. There was a specific relationship between distance, number of pixels in the image, and area of the image:

$$D = \beta \sqrt{\frac{S}{x}}$$

(6)

where β was a constant collectively composed by F, S, and ω.

2.2. Model Establishment

The maximum projected area of large fruits such as pomelo and even larger fruits would range from 100 cm^2 to 324 cm^2 since the pomelo longitudinal axis length ranged over 100~170 mm mostly. Therefore, the square and circular checkerboard calibration plates, such as those in Figure 3, with the areas of 100 cm^2, 121 cm^2, 144 cm^2, 169 cm^2, 196 cm^2, 225 cm^2, 256 cm^2, 289 cm^2, and 324 cm^2 as 9 groups were assigned and prepared for the establishment of models.

Under the condition that there was no sight line blocking within a certain positive range, as shown in Figure 4. The checkerboard was fixed, ensuring the checkerboard plane was perpendicular to the horizontal plane. For each checkerboard, 51 images were shot in the distance range from 0.25 m to 1.5 m at an interval of 0.025 m. Totally, 51 images shot from identical checkerboards were set as one group.

The number of pixels in the target image was calculated by Photoshop's lasso tool. Then the information about square checkerboards and circular checkerboards (distance, area, and number of pixels) was loaded into MATLAB for two specific models whose formations were similar to Equation (6). The specification of the computer used was a LAPTOP-5V7TD45J with a CPU of AMD Ryzen 5 3500U with Radeon Vega Mobile Gfx (2.10 GHz).
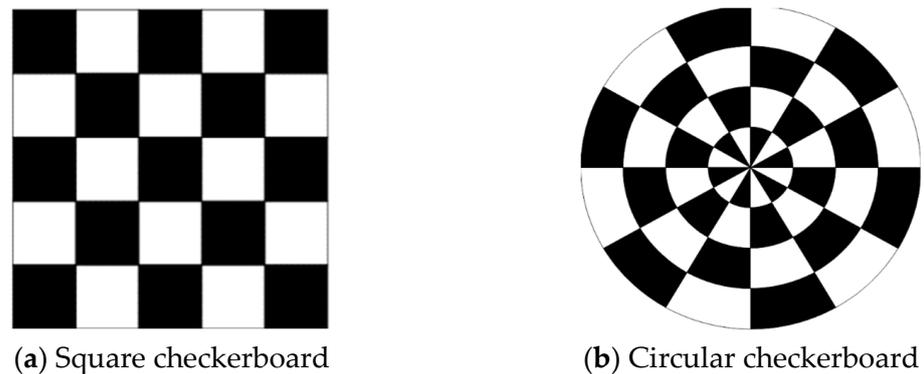
(**a**) Square checkerboard   (**b**) Circular checkerboard

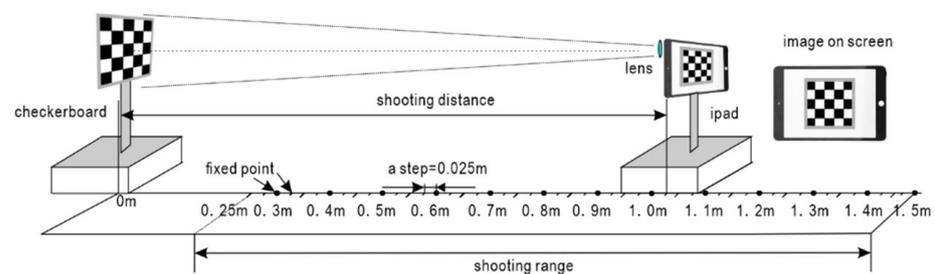**Figure 3.** Two patterns of the checkerboard.

Figure 4. The shooting device and the site environment for shooting.

In addition, to verify and polish the models, watermelon, and pomelo, two of the typical large-sized fruit, were selected to collect the image data under the same conditions. 12 groups of images were gathered together, involving 4 oval watermelons, 4 pyriform pomelos, and 4 oblate pomelos.

Each sample was assigned to pick four sides out for testing at an interval of 90°, as shown in Figure 5. The number of pixels and distance were recorded automatically. However, due to the irregular fruit's outline, the maximum projected area of the fruit target could not be measured directly. Therefore, an estimation approach to the area was applied, considering there was an obvious linear relationship between area and number of pixels, in particular when the camera was far enough away. For square checkerboards and circular checkerboards, the relationships between the number of pixels and the grid area were shown in Figure 6, respectively, which could acquire estimated areas of the fruit. Further, all the results of the maximum projected area of samples calculated by two different methods are listed in Table 1. The predicted distance could be obtained when the number of pixels and the projected area of fruits were input into the models.
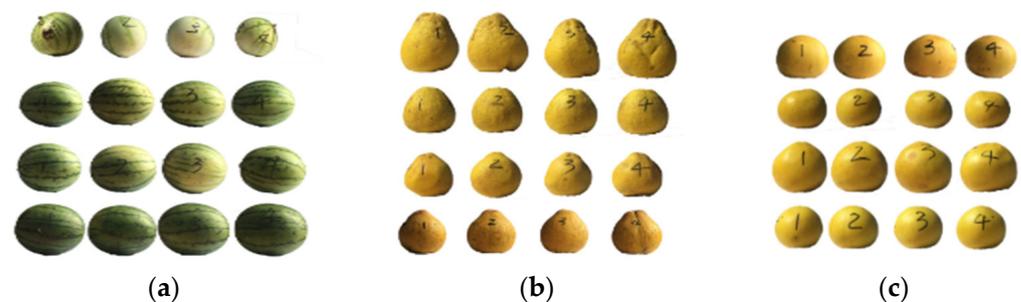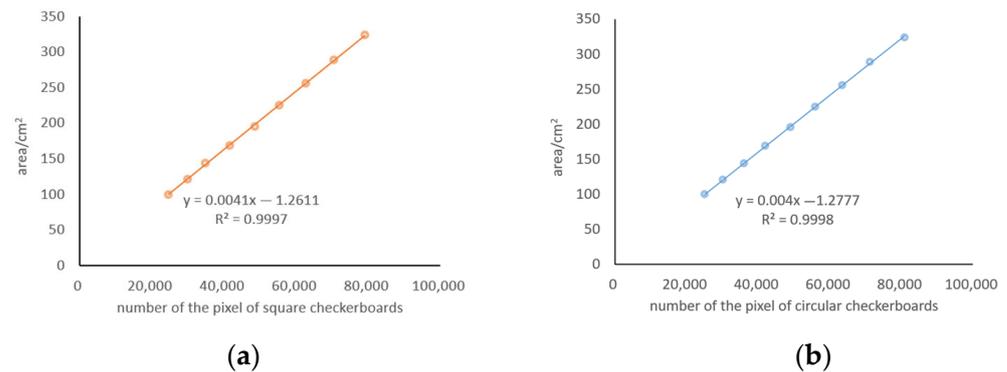

(**a**)   (**b**)   (**c**)

Figure 5. All samples of large fruits and their own 4 test sides: (**a**) oval watermelons; (**b**) pyriform pomelos; (**c**) oblate pomelos.

(a)

(b)

Figure 6. The linear relationships between number of pixels and the area of checkerboards included (**a**) square and (**b**) circular.

Table 1. The information was calculated by the area and number of pixels of sample fruit—type, shape, and number—as well as the maximum projected area of the four test sides obtained.

Fruit	Shape	Sample Number	Maximum Projected Area/cm²							
			1st Side		2nd Side		3rd Side		4th Side	
			Square	Circular	Square	Circular	Square	Circular	Square	Circular
Watermelon	Oval	1	110.72	107.97	100.10	97.61	105.47	102.85	98.09	95.65
Watermelon	Oval	2	145.98	142.37	153.09	149.31	150.29	146.58	150.17	146.46
Watermelon	Oval	3	176.27	171.93	166.29	162.18	178.96	174.55	168.04	163.89
Watermelon	Oval	4	206.83	201.74	201.52	196.56	209.09	203.94	204.83	199.79
Pomelo	Pyriform	1	172.54	168.29	178.82	174.41	169.55	165.37	176.97	172.61
Pomelo	Pyriform	2	124.06	120.99	125.84	122.73	124.74	121.65	126.37	123.24
Pomelo	Pyriform	3	101.30	98.78	103.65	101.07	102.74	100.19	102.98	100.42
Pomelo	Pyriform	4	128.40	125.22	128.75	125.56	124.32	121.23	125.19	125.19
Pomelo	Oblate	1	79.19	77.21	76.35	74.44	77.72	75.78	76.22	74.31
Pomelo	Oblate	2	137.68	134.28	129.98	126.76	136.33	132.96	129.16	125.96
Pomelo	Oblate	3	133.42	130.12	139.58	136.13	136.81	133.42	139.71	136.26
Pomelo	Oblate	4	125.22	123.48	125.56	123.79	121.23	118.95	125.19	122.68

In this work, the accuracy of the model was classified into 3 grades to define the precise degree of prediction according to the number of relative errors—less than 5% relative error was assessed as "High" accuracy, between 5% and 10% was regarded as "Middle" accuracy, and higher than 10% was rated as "Low" accuracy. The model's exact predicting ability would be justified by counting up the number of every grade. Furthermore, the maximum relative error and absolute error of the two prediction models were also essential parameters for evaluating the superiority of the models.

2.3. Model Error Modification

Considering the shape of the fruit is always irregular, errors in predicting results are inevitable. A modification for that, as an essential method, could effectively enhance the accuracy of a model. After the specific change tendency between the relative error and real distance was analyzed, a new modification polynomial was obtained. Via modification, the accuracy of models would be compared and evaluated again to observe the change.

3. Results and Discussion

3.1. Establishment of Models

According to the 18 groups of image data of the checkerboards with two shapes, a scatter diagram—Figure 7a—belonged to Model S, which described vividly specific distribution conditions in a three-dimensional coordinate system where the *x* axis, *y* axis, and *z* axis represented respectively the number of pixels, area of the checkerboard, and shooting distance. The curve-fitting surface of Figure 7b, whose distance starts at 0.25 m and

culminates at 1.50 m in a three-dimensional coordinate, was obtained. The two functional relationships between Model S and Model C were:

$$d_s = 22.64 \cdot S^{0.4936} \cdot x^{-0.4932} \tag{7}$$

and

$$d_c = 23.51 \cdot S^{0.4949} \cdot x^{-0.4990} \tag{8}$$

d_s and d_c, respectively, represented the predicting distances of models S and C; S was the maximum projected area of the fruit sample, and x was the number of pixels in the image of the fruit sample. Table 2 lists the number of pixels and the image distance of a square-shaped checkerboard and a circular-shaped checkerboard with a 225 cm^2 area.

Table 2. The number of pixels and shooting distance of 225 cm^2 square checkerboard and circular checkerboard.

| Distance (*d*)/m | Number of Pixels (*x*) | | Distance (*d*)/m | Number of Pixels (*x*) | |
	Square Checkerboard	Circular Checkerboard		Square Checkerboard	Circular Checkerboard
0.250	2,125,400	2,027,978	0.900	157,169	152,677
0.275	1,810,936	1,662,104	0.925	145,664	144,729
0.300	1,458,858	1,386,520	0.950	137,918	138,808
0.325	1,277,878	1,199,561	0.975	133,377	131,268
0.350	1,044,969	1,018,685	1.000	125,000	124,805
0.375	941,468	888,487	1.025	118,917	119,939
0.400	799,591	776,853	1.050	112,923	112,443
0.425	721,767	688,263	1.075	108,788	108,577
0.450	628,723	619,131	1.100	102,864	104,056
0.475	575,948	555,015	1.125	97,706	99,427
0.500	502,876	500,698	1.150	93,544	95,202
0.525	471,042	454,438	1.175	90,278	90,870
0.550	413,690	414,606	1.200	86,200	86,456
0.575	385,746	377,284	1.225	83,074	84,213
0.600	346,711	347,165	1.250	79,721	76,901
0.625	325,720	322,823	1.275	77,675	74,469
0.650	294,896	297,677	1.300	73,686	73,686
0.675	275,735	274,991	1.325	71,471	71,481
0.700	254,463	253,468	1.350	68,586	69,220
0.725	239,976	235,587	1.375	65,719	66,114
0.750	221,816	221,410	1.400	62,926	64,467
0.775	209,194	206,859	1.425	61,690	61,875
0.800	193,000	194,518	1.450	58,614	59,503
0.825	183,087	183,071	1.475	56,673	57,292
0.850	170,805	172,736	1.500	55,230	55,939
0.875	163,039	164,536			

After the models S and C were established, tests were designed to verify the accuracy of the model predictions. Two watermelons, two pyriform pomelos, and two oblate pomelos were picked, and each one gave four different aspects for camera shooting. The maximum projected area would be calculated through the linear relationship between the number of pixels and the area of checkerboards described in Figure 4. Hence, the predicting distance could be obtained when the maximum projected area and number of pixels are known inputs to models. The following Table 3 is an illustration of predicting results when one of the aspects of a watermelon was 110.72 cm^2 in area.
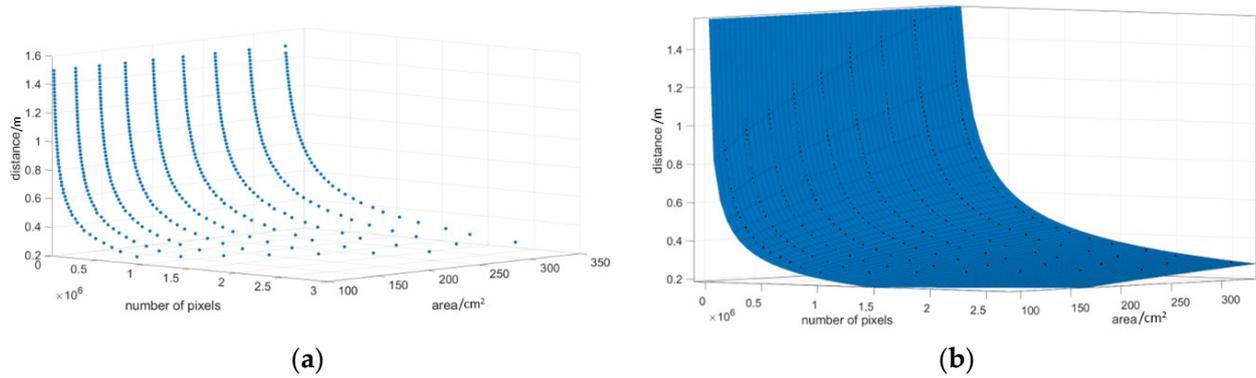
Figure 7. The functional relationship between the area, number of pixels, and distance, including (**a**) the scatter point diagrams of Model S and (**b**) the curve surfaces from the scatter diagram fitting of Model S.

Table 3. Testing results for model S and model C with the watermelon of 110.72 cm^2 maximum projected area.

d/m	Model S			Model C			d/m	Model S			Model C		
	d_{ps}/m	e_{as}/m	e_{rs}/%	d_{pc}/m	e_{ac}/m	e_{rc}/%		d_{ps}/m	e_{as}/m	e_{rs}/%	d_{pc}/m	e_{ac}/m	e_{rc}/%
0.250	0.287	0.0375	14.994	0.276	0.0262	10.497	0.900	0.912	0.012	1.289	0.887	0.013	1.435
0.275	0.314	0.039	14.214	0.302	0.027	9.853	0.925	0.944	0.019	2.013	0.919	0.006	0.694
0.300	0.338	0.038	12.525	0.325	0.025	8.315	0.950	0.967	0.017	1.834	0.942	0.008	0.840
0.325	0.361	0.036	11.156	0.348	0.023	7.076	0.975	0.989	0.014	1.436	0.963	0.012	1.205
0.350	0.383	0.033	9.481	0.369	0.019	5.531	1.000	1.015	0.015	1.471	0.989	0.011	1.142
0.375	0.411	0.036	9.569	0.396	0.021	5.696	1.025	1.041	0.016	1.574	1.015	0.010	1.015
0.400	0.435	0.035	8.742	0.420	0.020	4.964	1.050	1.063	0.013	1.277	1.037	0.013	1.281
0.425	0.461	0.036	8.371	0.445	0.020	4.671	1.075	1.089	0.014	1.284	1.062	0.013	1.249
0.450	0.483	0.033	7.408	0.467	0.017	3.796	1.100	1.112	0.012	1.073	1.084	0.016	1.432
0.475	0.506	0.031	6.445	0.489	0.014	2.916	1.125	1.140	0.015	1.348	1.112	0.013	1.136
0.500	0.528	0.028	5.626	0.511	0.011	2.172	1.150	1.159	0.009	0.816	1.131	0.019	1.638
0.525	0.553	0.028	5.326	0.535	0.010	1.933	1.175	1.183	0.008	0.695	1.155	0.020	1.734
0.550	0.578	0.028	5.048	0.559	0.009	1.713	1.200	1.213	0.013	1.077	1.184	0.016	1.334
0.575	0.600	0.025	4.390	0.581	0.006	1.119	1.225	1.234	0.009	0.742	1.205	0.020	1.642
0.600	0.624	0.024	3.937	0.604	0.004	0.722	1.200	1.256	0.006	0.515	1.227	0.023	1.845
0.625	0.648	0.023	3.679	0.628	0.003	0.514	1.275	1.284	0.009	0.685	1.254	0.021	1.656
0.650	0.673	0.023	3.601	0.653	0.003	0.481	1.300	1.308	0.008	0.616	1.278	0.022	1.703
0.675	0.698	0.023	3.445	0.677	0.002	0.370	1.325	1.329	0.004	0.290	1.298	0.027	2.005
0.700	0.722	0.022	3.196	0.701	0.001	0.165	1.350	1.351	0.001	0.107	1.321	0.029	2.165
0.725	0.749	0.024	3.247	0.727	0.002	0.253	1.375	1.378	0.003	0.195	1.347	0.028	2.059
0.750	0.770	0.020	2.633	0.748	0.002	0.312	1.400	1.415	0.015	1.061	1.383	0.017	1.183
0.775	0.791	0.016	2.041	0.768	0.007	0.858	1.425	1.423	0.002	0.166	1.391	0.034	2.378
0.800	0.817	0.017	2.119	0.794	0.006	0.747	1.450	1.444	0.006	0.417	1.412	0.038	2.607
0.825	0.841	0.016	1.983	0.818	0.007	0.847	1.475	1.471	0.004	0.272	1.439	0.036	2.445
0.850	0.865	0.015	1.798	0.842	0.008	0.997	1.500	1.490	0.010	0.640	1.458	0.042	2.790
0.875	0.891	0.016	1.835	0.867	0.008	0.929							

Note: d—distance; S—the maximum projected area; d_{ps}, d_{pc}—predicting distance of square model and circular model; e_{as}, e_{ac}—absolute error of two models; e_{rs}, e_{rc}—relative error of two models.

The comparison of average absolute error, absolute error maximum, average relative error, and relative error maximum was exhibited in Table 4. It showed that model C had more priorities over model S for whatever shape: model C had a prominent advantage over accuracy with a 42 average number of "High" accuracy compared with model S at 35; while, for "Middle" and "Low" accuracy, model S obviously had more shares, taking account of an average of 8.75 and 7.25. Hence, it could be concluded that model C had more accuracy in predicting the imaging distance of the watermelon and pomelo targets.

Table 4. The error extremums and accuracy levels of two models.

| Fruit	Model S				Model C			
	Average Absolute Error/m	Absolute Error Maximum/m	Average Relative Error/%	Relative Error Maximum/%	Average Absolute Error/m	Absolute Error Maximum/m	Average Relative Error/%	Relative Error Maximum/%
W1	0.024	0.065	4.370	26.110	0.020	0.053	3.216	21.355
W2	0.022	0.048	4.167	19.081	0.020	0.048	3.022	14.337
OP1	0.017	0.037	2.080	14.557	0.018	0.043	2.574	11.614
OP2	0.020	0.051	3.282	20.276	0.019	0.042	3.117	16.997
PP1	0.027	0.060	4.991	23.855	0.020	0.047	3.380	18.930
PP2	0.023	0.051	4.360	20.483	0.020	0.042	3.171	15.778
RH		68.63%				82.35%		
RM		17.17%				11.27%		
RL		14.21%				6.37%		

Note: W1—watermelon NO.1, W2—watermelon NO.2; OP1—oblate pomelo NO.1, OP2—oblate pomelo NO.2; PP1—pyriform pomelo NO.1, PP2—pyriform pomelo NO.2. RH—rate of "High" accuracy; RM—rate of "Middle" accuracy; RL—rate of "Low" accuracy.

3.2. Model Modification

The detailed distribution of the predicted results was described in Figure 8, illustrating that the results would fluctuate around the ideal conditions with an error gap. In Figure 9, Range 1, Range 2, and Range 3 were respectively defined as the distance range where the predicting results could not be kept within [−10%,10%], the distance range where the predicting results could be kept within [−10%,10%] but some were over 5%, and the distance range where the predicting results can be curbed within [−5%,5%]. Based on Figure 9, which illustrated the conditions of relative error, three ranges, including Range 1 (0.25~0.525 m), Range 2 (0.525~0.8 m), and Range 3 (0.8~1.5 m), were divided. When shooting distance falls in Range 1, the results could not remain reliable; if in Range 2, the results could get an accuracy assurance within −10~10% error; if in Range 3, then the results would be accurate with only −5~5% error. Hence, when the distance is between 0.8 m and 1.5 m, it would make a sound effect; if the distance falls between 0.8 m and 1.5 m, there is a large improvement room to be more accurate for prediction.
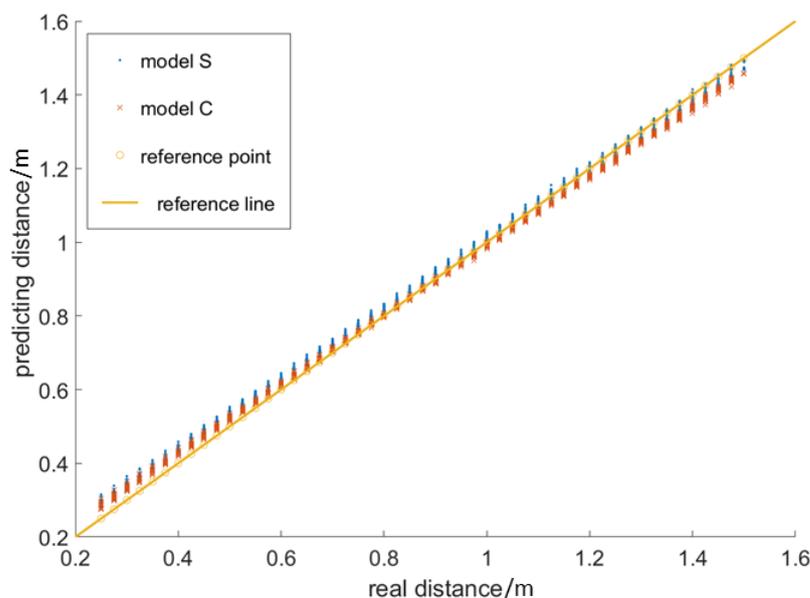
Figure 8. A series of points in a cartesian plane coordinate system with real distance as a horizontal axis and predicting distance as vertical coordinates for two models where the reference point and reference line are ideal goals.
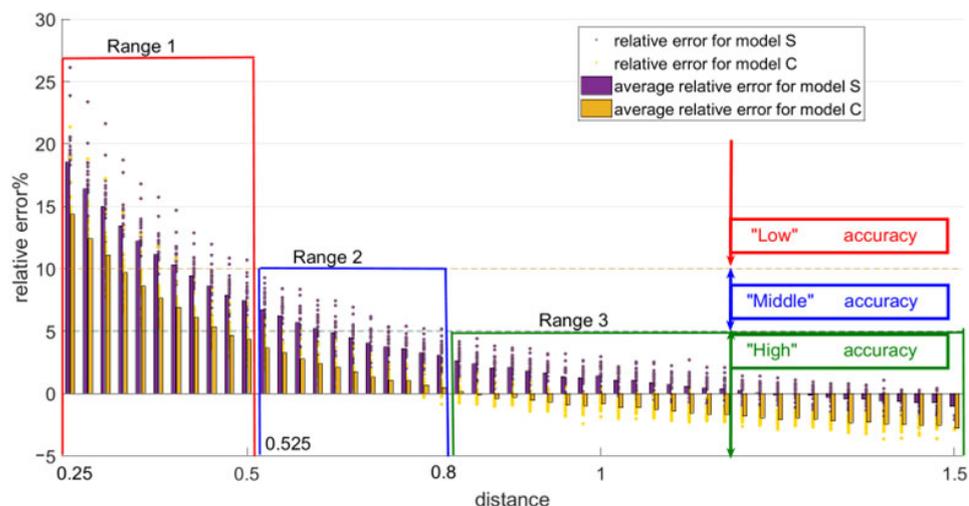
Figure 9. Distribution position of relative predicting distance error of model S and model C around distance, the average relative error, which are represented by bars at every distance.

According to Figure 10, where the fitting lines of two models were illustrated, the predicted error value changed step by step from negative to positive when the distance slowly decreased from 1.5 m to 0.25 m. Such a rough tendency for error fluctuation would be regarded as a ground for modifying the form of models. Namely, an approximately linear relationship would exist between real distance and predicting distance for two models. Via repeated fittings, two functional expressions could be described as:

$$d_{\mathrm{ps}} = 0.95274 d_{\mathrm{rs}} + 0.06 \tag{9}$$

and

$$d_{\mathrm{pc}} = 0.93912 d_{\mathrm{rc}} + 0.05 \tag{10}$$

d_{ps} and d_{pc} represented predicting distance for model S and model C, and d_{rs} and d_{rc} represented real distance for model S and model C.
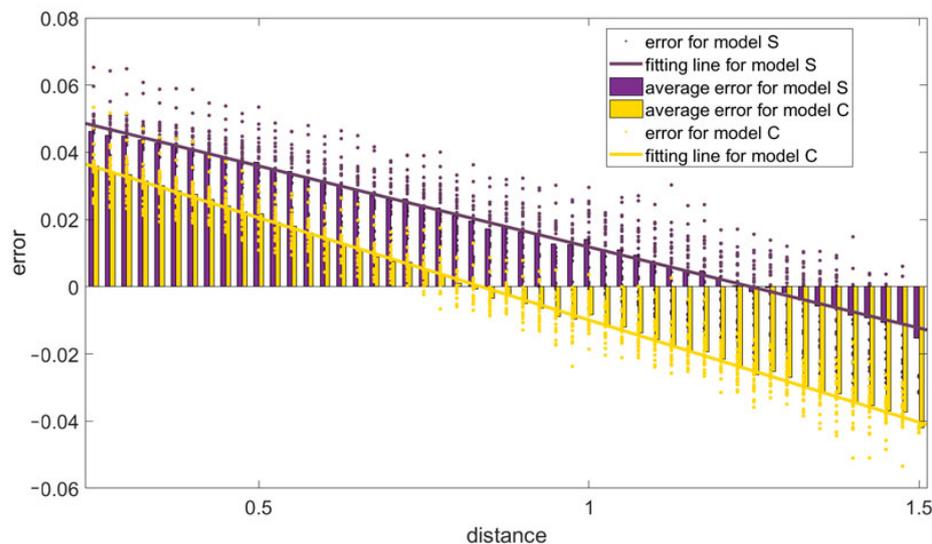
Figure 10. Distribution position of predicting distance error of model S and model C around distance, where the bars represent the average error at every distance.

Substituting the above Equations (9) and (10) into Equations (7) and (8) and modifying models, respectively, could be:

$$d_s = 23.83 \cdot S^{0.4936} \cdot x^{-0.4932} - 0.063 \tag{11}$$

and

$$d_c = 25.01 \cdot S^{0.4949} \cdot x^{-0.499} - 0.054 \tag{12}$$

A comprehensive analysis, referring to several potential factors for predicting results, including average absolute error, absolute error maximum, average relative error, and relative error maximum, as well as accuracy shares, was carried out. From Table 5 and Figure 11, a clear improvement in accuracy for prediction could be seen after a simple modification. On the one hand, the number of "High" rating shares had a huge increase from 68.63% to 98.04% for Model S and from 82.35% to 98.53% for Model C. It meant almost all prediction results had less than 5% absolute error. On the other hand, the extinction of "Low" rating shares dropped significantly from 14.21% to 0 for Model S and from 6.73.% to 0 for Model C. Compared with model S, model C still retained a comprehensive edge in accuracy, with a higher average number of "High" classes and a lower absolute error maximum. Via modification, two methods had made the "High" accuracy rate increase to more than 98%.

**Table 5.** The error extremums of the modified models.

Fruit	Model S				Model C			
	Average Absolute Error	Absolute Error Maximum	Average Relative Error%	Relative Error Maximum%	Average Absolute Error	Absolute Error Maximum	Average Relative Error%	Relative Error Maximum%
W3	0.009	0.018	0.929	4.519	0.005	0.013	0.809	4.625
W4	0.012	0.020	1.183	6.609	0.007	0.018	0.811	6.622
OP3	0.003	0.011	0.276	4.239	0.003	0.012	0.818	3.772
OP4	0.007	0.024	0.681	2.591	0.003	0.012	0.821	2.399
PP3	0.004	0.015	0.387	3.767	0.003	0.012	0.826	3.425
PP4	0.006	0.018	0.610	4.166	0.007	0.021	0.830	3.850
RH	98.04%				98.53%			
RM	1.96%				1.47%			
RL	0				0			

Note: W3—watermelon NO.3, W4—watermelon NO.4; OP3—oblate pomelo NO.3, OP4—oblate pomelo NO.4; PP3—pyriform pomelo NO.3, PP4—pyriform pomelo NO.4. RH—rate of "High" accuracy; RM—rate of "Middle" accuracy; RL—rate of "Low" accuracy.
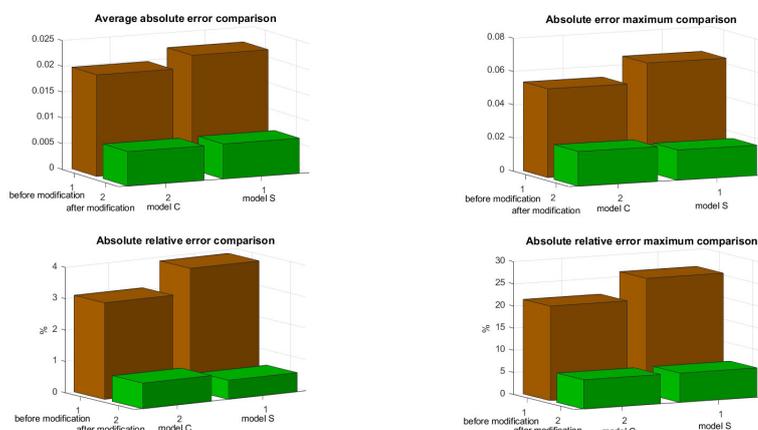
(**a**) Four error indexes: average absolute error, absolute error maximum, average relative error, and relative error maximum.

Figure 11. *Cont.*

(**b**) "High", "Middle", and "Low" accuracy shares.
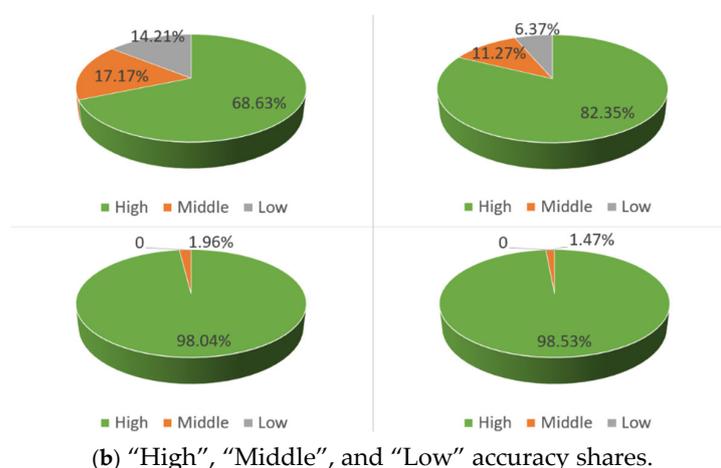
Figure 11. The compared results before and after modification.

The error of this model is mainly derived from two factors. The first one was the irregularity of pomelo and watermelon, which have an essential difference between an absolute square shape and a circle shape. Hence, the error was unavoidable in predicting the results. Another one was the transformation of the shoot side at different distances for even the same pomelo or watermelon fruit, due to the fact that when the camera is close to the fruit, the shoot side always tends to be smaller in view of the fruit as a three-dimensional object.

4. Conclusions

In this work, distance detection models based on the images of square checkerboards and circular checkerboards from only one camera were established, respectively. Two models described the special relationships among the three main imaging factors—shooting distance, the maximum projected area of the target, and number of pixels. The accuracy was tested on 2 oblate pomelos, 2 pyriform pomelos, and 2 oval watermelons. After error analysis, two models were modified and tested again. The main conclusions were as follows:

(1) The regression model based on a square-shaped checkerboard could be presented as Equation (7), while a circular-shaped checkerboard could be presented as Equation (8).
(2) According to the error analysis, two models were modified to Equations (11) and (12).
(3) Two models had over 98% "high" accuracy and no "low" accuracy. By comparison, the model of a circular checkerboard would be more predictively accurate than the one of a square checkerboard when they were used for large fruits.
(4) Two models showed a sound-predicting effect with a 0.8~1.5 m distance where the camera was far away shooting an object when applied to the watermelon and pomelo fruit, due to the relative error being controlled within 2%.

The results indicated that the proposed method provided a new way to achieve accurate distance measurement for fruit-picking devices. However, there would be some problems to overcome and deserve exploration in the future, such as the definite application and operation of that, the spreading of effective range, and further improvement of accuracy.

Author Contributions: Conceptualization, D.Z. and J.L.; formal analysis, D.Z., Y.L. and Y.W.; funding acquisition, J.L.; methodology, D.Z. and J.L.; project administration, J.L.; software, D.Z.; supervision, J.L.; validation, D.Z., Y.L., Y.W. and W.L.; visualization, D.Z.; writing—original draft, D.Z.; writing—review and editing, J.L. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kang, D.; Chen, Z.J.; Fan, Y.H.; Li, C.; Mi, C.; Tang, Y.H. Optimization on kinematic characteristics and lightweight of a camellia fruit picking machine based on the kriging surrogate model. *Mech. Ind.* **2021**, *22*, 16. [CrossRef]
2. Cao, X.; Yan, H.; Huang, Z.; Ai, S.; Xu, Y.; Fu, R.; Zou, X. A multi-objective particle swarm optimization for trajectory planning of fruit picking manipulator. *Agronomy* **2021**, *11*, 2286. [CrossRef]
3. Pan, S.; Ahamed, T. Pear Recognition in an Orchard from 3D stereo camera datasets to develop a fruit picking mechanism using mask R-CNN. *Sensors* **2022**, *22*, 4187. [CrossRef] [PubMed]
4. Rehman, A.; Saba, T.; Kashif, M.; Fati, S.M.; Bahaj, S.A.; Chaudhry, H. A revisit of Internet of Things technologies for monitoring and control strategies in smart agriculture. *Agronomy* **2022**, *12*, 127. [CrossRef]
5. Kang, H.W.; Chen, C. Fruit Detection and Segmentation for Apple Harvesting Using Visual Sensor in Orchards. *Sensors* **2019**, *19*, 4599. [CrossRef] [PubMed]
6. Zhang, F.; Chen, Z.; Wang, Y.; Bao, R.; Chen, X.; Fu, S.; Tian, M.; Zhang, Y. Research on flexible end-effectors with humanoid grasp function for small spherical fruit picking. *Agriculture* **2023**, *13*, 123. [CrossRef]
7. Shi, Y.; Zhang, W. A "global-local" visual servo system for picking manipulators. *Sensors* **2020**, *20*, 3366. [CrossRef] [PubMed]
8. Han, Y.; Routray, A.; Adeghate, J.O.; MacLachlan, R.A.; Martel, J.N.; Riviere, C.N. Monocular vision-based retinal membrane peeling with a handheld robot. *ASME J. Med. Devices* **2021**, *15*, 031014. [CrossRef] [PubMed]
9. Fonder, M.; Ernst, D.; Van Droogenbroeck, M. M4depth: A motion-based approach for monocular depth estimation on video sequences. *arXiv* **2021**, arXiv:2105.09847. [CrossRef]
10. Tao, Y.; Zhou, J. Automatic apple recognition based on the fusion of color and 3D feature for robotic fruit picking. *Comput. Electron. Agric.* **2017**, *142*, 388–396. [CrossRef]
11. Xiong, J.; Lin, R. The recognition of litchi clusters and the calculation of picking point in a nocturnal natural environment. *Biosyst. Eng.* **2018**, *166*, 44–57. [CrossRef]
12. Wang, C.; Li, Z. Weed recognition using SVM model with fusion height and monocular image features. *Trans. Chin. Soc. Agric. Eng.* **2016**, *32*, 165–174.
13. Liu, X.; Chen, S.W. Monocular camera based fruit counting and mapping with semantic data association. *IEEE Robot. Autom. Lett.* **2019**, *4*, 2296–2303. [CrossRef]
14. Roberts, L. Machine Perception of Three-Dimensional Solids. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 1963.
15. Yamaguti, N.; Oe, S. A method of distance measurement by using monocular camera. In Proceedings of the 36th SICE Annual Conference, International Session Papers, Tokushima, Japan, 29–31 July 1997; pp. 1255–1260.
16. Kim, H.; Lin, C.S. Distance measurement using a single camera with a rotating mirror. *Int. J. Control. Autom.* **2005**, *3*, 542–551.
17. Alizadeh, P.; Zeinali, M. A real-time object distance measurement using a monocular camera. In Proceedings of the Iasted International Conference on Modelling Simulation & Optimatization, Banff, AB, Canada, 1 January 2013; pp. 237–242.
18. Xiao, D.; Zhai, J. Target distance measurement method with monocular vision for wheeled mobile robot. *Comput. Eng.* **2017**, *43*, 287–291.
19. Huang, L.; Chen, Y. Measurement the absolute distance of a front vehicle from an in-car camera based on monocular vision and instance segmentation. *J. Electron. Imaging* **2018**, *27*, 043019.1–043019.10. [CrossRef]
20. Bazrafkan, S.; Javidnia, H. Semiparallel deep neural network hybrid architecture: First application on depth from monocular camera. *J. Electron. Imaging* **2018**, *27*, 19. [CrossRef]
21. Wang, X.; Zhou, B. Recognition and distance estimation of an irregular object in package sorting line based on monocular vision. *Int. J. Adv. Robot. Syst.* **2019**, *16*, 1729881419827215. [CrossRef]
22. Bianco, S.; Buzzelli, M. A unifying representation for pixel-precise distance estimation. *Multimed. Tools Appl.* **2019**, *78*, 13767–13786. [CrossRef]