applied sciences

MDPI

Article

Distance Estimation with a Stereo Camera and Accuracy Determination

Arnold Zaremba [1] and Szymon Nitkiewicz [1,2,*]

[1] Department of Mechatronics, University of Warmia and Mazury in Olsztyn, 10-719 Olsztyn, Poland; arnold.zaremba@uwm.edu.pl
[2] Department of Neurosurgery, Collegium Medicum, University of Warmia and Mazury in Olsztyn, 10-719 Olsztyn, Poland
* Correspondence: szymon.nitkiewicz@uwm.edu.pl

Featured Application: This research addresses the critical problem of accurate distance estimation using stereo camera systems, which are increasingly relevant for applications in metrology, robotics, and automated systems. The manuscript provides theoretical insights and practical experimental results on the performance and accuracy of stereo vision for distance estimation.

Abstract: Distance measurement plays a key role in many fields of science and technology, including robotics, civil engineering, and navigation systems. This paper focuses on analyzing the precision of a measurement system using stereo camera distance measurement technology in the context of measuring two objects of different sizes. The first part of the paper presents key information about stereoscopy, followed by a discussion of the process of building a measuring station. The Mask R-CNN algorithm, which is a deep learning model that combines object detection and instance segmentation, was used to identify objects in the images. In the following section, the calibration process of the system and the distance calculation method are presented. The purpose of the study was to determine the precision of the measurement system and to identify the distance ranges where the measurements are most precise. Measurements were made in the range of 20 to 70 cm. The system demonstrated a relative error of 0.95% for larger objects and 1.46% for smaller objects at optimal distances. A detailed analysis showed that for larger objects, the system exhibited higher precision over a wider range of distances, while for smaller objects, the highest accuracy was achieved over a more limited range. These results provide valuable information on the capabilities and limitations of the measurement system used, while pointing out directions for its further optimization.

Keywords: distance measurement; stereo vision; image processing; accuracy determination

Citation: Zaremba, A.; Nitkiewicz, S. Distance Estimation with a Stereo Camera and Accuracy Determination. *Appl. Sci.* **2024**, *14*, 11444. https://doi.org/10.3390/app142311444

Academic Editor: Samuel Morillas

Received: 18 October 2024
Revised: 20 November 2024
Accepted: 2 December 2024
Published: 9 December 2024

1. Introduction

Distance estimation using stereo cameras is one of the key issues in the field of computer vision and robotics, which finds application in various areas of technology and industry. The technology is based on the use of two cameras placed at a certain distance from each other, which capture images of the same scene from different viewing angles. The basic concept behind this method is that the difference between the images captured by the two cameras, known as binocular disparity (binocular parallax), contains information about the depth of the scene. It is the main depth cue that makes stereoscopic images appear three-dimensional. However, in many scenarios, the range of depth that can be reproduced by this signal is severely limited and is usually fixed due to limitations imposed by the displays. For example, due to the low angular resolution of current automultiscope displays, they can only reproduce a small depth range. Analyzing these differences, which is the process of stereoscopy, makes it possible to recreate the three-dimensional structure of the environment and precisely determine the distance to individual objects in the cameras' field of view.

This technology is widely used in many fields. In mobile robotics, for example, stereo systems are used for navigation and obstacle avoidance, enabling robots to navigate in unfamiliar and dynamic environments [1–3]. In the automotive industry, stereo cameras are an integral part of advanced driver assistance systems that enhance road safety through automatic braking, lane keeping, or pedestrian recognition. Also in autonomous vehicles, there are systems to measure the distance between vehicles for autonomous driving based on processing images obtained from stereo cameras [1,2]. Stereo cameras are also used to monitor the condition of road surfaces, including the detection of ice, water, snow, and dry asphalt. The system is based on the analysis of changes in the polarization of light reflected from the road surface, and the recognition accuracy has been improved through texture analysis, which evaluates the image contrast. Although tests have shown the need for further improvements, especially to adapt the system to changing lighting conditions, the technology is a promising solution to detect road hazards [3].

There are studies of distance measurement using a stereo camera, in which the test subject was a human. The process included steps such as camera calibration, image rectification, disparity calculation, and 3D reconstruction. The distance of the subject was measured using Euclid's distance measurement method to determine the shortest distance from the center of the bounding box to both cameras [4]. Stereoscopic cameras have also been used to develop an inexpensive diagnostic system for ophthalmology that could find application in developing countries. Due to its stereoscopic capabilities and low manufacturing costs, the prototype camera provides clear images of the fundus of the eye, allowing the user to choose between photo and video. This innovative system represents a significant step forward in available medical technology, meeting the needs of professionals working in resource-limited environments [5]. Stereo cameras have also been used to visualize 3D ultrasound data of the breast to better understand the depth and shape of tumors. For this purpose, an acquisition system with automatic ultrasound head travel has been developed that creates a series of parallel image cross sections and then combines them into a 3D model. With a stereoscopic monitor, doctors can view these data in the form of spatial projections, improving the accuracy of cancer diagnosis [6].

To measure the distance of an object from the stereo cameras, it is first necessary to accurately detect the object in the images captured by both cameras. To do this, advanced computer vision techniques, such as Mask R-CNN (Mask Region-Based Convolutional Neural Network), can be used. Mask R-CNN is a powerful tool for segmenting objects in images that not only identifies and classifies objects, but also accurately determines their contours by generating segmentation masks [7–9]. This makes it possible to isolate specific objects from the background, which is crucial for the accurate matching of correspondence points between images from the two cameras. The use of Mask R-CNN in the process of distance estimation not only allows for the improvement of the precision of object detection, but also significantly improves the entire process of 3D scene reconstruction, leading to more reliable and accurate distance measurements.

Another advanced method that can be used to measure depth is the Spiking U-Net architecture, based on spiking neural networks (SNNs). These networks differ from traditional deep neural networks in that they mimic the action of biological neurons, processing information in the form of pulses, known as 'spikes', rather than continuous signals. In the U-Net architecture, originally designed for the task of segmenting medical images, such spikes are used to more efficiently detect the edges and contours of objects in three-dimensional scenes. Spiking U-Net generates detailed segmentation masks, which in turn help with precise depth reconstruction. Spiking U-Net is an interesting alternative to classic networks such as Mask R-CNN, especially in the context of real-time applications. The use of this architecture can significantly improve precision in depth measurements, while enabling the more economical use of computing resources, which is important in the context of mobile or portable measurement systems [10,11].

It was decided to conduct research to accurately determine the precision of the measurement system and identify specific distance ranges where the results obtained have the

highest accuracy. Such a research goal allows for a better understanding of the capabilities and limitations of the systems used and helps determine the optimal operating conditions relevant to future practical applications of this technology.

2. Materials and Methods

To carry out the test, a simple test bench was set up, as illustrated in Figure 1. The main component of this bench was ESP32 microcontrollers (Espressif Systems, SSE: 688018.SH) with an attached OV2640 camera module (Waveshare Shenzhen, China), which offers a maximum resolution of up to 1600 × 1200 pixels [12,13]. The measurement setup consisted of two ESP32-CAM (OV2640) modules mounted in parallel on a single plane at a fixed distance of 5.4 cm from each other to facilitate stereovision, in which two cameras, acting as a stereo pair, captured images simultaneously. Precise parallel alignment and a fixed distance between the cameras were crucial to accurate depth measurement and triangulation. The cameras were securely mounted to avoid any displacement during the measurement process. A measuring tape was placed in front of the cameras, marking distances from 20 to 70 cm in 2.5 cm increments. This made it possible to precisely control the distance measurements. Three measurements were taken for each marked distance to ensure the precision and repeatability of the data. The cameras were mounted in the same perfectly even plane and precisely aligned with the edges to minimize the risk of misalignment. Minor deviations in the camera axes could introduce errors in the triangulation process, which would directly affect the precision of depth measurements. Even slight misalignment causes discrepancies in correspondence points between the images from the two cameras, which in turn leads to errors in distance determination.

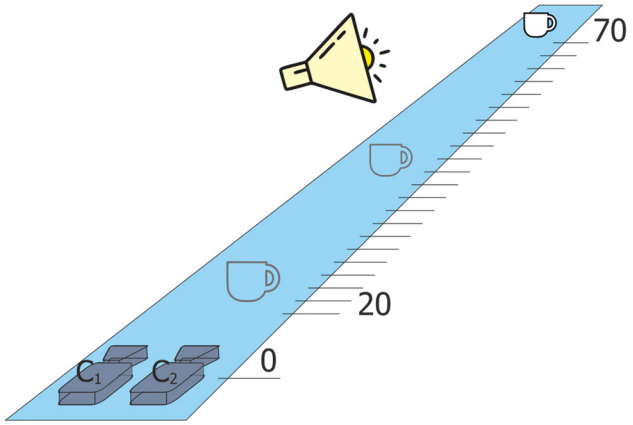

Figure 1. Schematic view of the test bed for distance measurement with stereo cameras.

This project used Mask R-CNN to detect and segment objects in images from two ESP32-CAM cameras. Mask R-CNN, which is based on the ResNet-50 architecture with Feature Pyramid Network (FPN) functionality, allows for precise object position determination and accurate segmentation. Images are loaded and processed in a tensor, which is then passed to a model that returns bounding boxes, class labels, confidence values, and segmentation masks [14–16]. The detected bounding boxes of the two images are compared to calculate the distances between the corresponding boxes. Functions that convert box coordinates to different representations, such as midpoints or corners, are used for accurate calculations. A cost matrix is created, taking into account vertical and horizontal offsets and area differences, to determine the optimal assignments between objects. A linear assignment method is used to determine which boxes from one image correspond to boxes from another, enabling the tracking of object movement. The results of detection and assignments are visualized in the images using box drawing, class labels, and segmentation masks. This makes it easy to verify the accuracy of the detection and assignments, which is crucial for further analysis and applications. The use of Mask R-CNN

in this project allowed for precise information about the movement and position of objects in space, which is essential for triangulation and object tracking.

To calibrate the cameras in the project, the CameraWebServer program, available from Arduino.IDE (1.8.19.) software as a sample program after installing the camera library, was first uploaded to the ESP32-CAM modules. This program allows access to the video stream from the cameras via a web interface, allowing easy monitoring and recording of live images. With CameraWebServer, it is also possible to adjust camera parameters such as brightness, contrast, white balance, and exposure, which is key to obtaining the higher-quality images needed for calibration. After running the program and obtaining a stable connection to the cameras, the calibration process proceeded, which involved taking images from both cameras at 30 and 50 cm distances. These images were then processed with a Python script that uses the OpenCV library and the Mask R-CNN model for object detection. The script starts by loading the images and preprocessing them. It then uses the aforementioned neural network model to detect and classify objects in the images. Once the objects are detected, the script calculates various parameters, such as the positions of the objects' centers of mass, the distances between the objects, and the differences in their surfaces. The next step is to calculate the cost of moving objects between two images. In the context of image processing, especially in stereovision, these costs take into account differences in horizontal positions (X-axis displacement), vertical positions (Y-axis displacement), and changes in object surfaces. These costs are crucial to correctly matching objects detected in images from the two cameras, as they make it possible to determine exactly which elements in one image correspond to elements in the other image. The process involves analyzing the differences between the images and finding the best match for each object, which is necessary to calculate the distances to those objects. The script matches objects between images and then calculates the distances to those objects, taking into account the known distance between cameras and viewing angles. An important aspect of this process is the calculation of the focal length and angle tangent, which are necessary to accurately calculate the final distance to the objects. The focal length is a parameter that describes the ability of a lens to focus light, which is crucial in determining image scaling. The angle tangent, on the other hand, refers to the ratio of the distance between cameras to the distance to the object in the scene. The final distance away is calculated on the basis of these parameters, allowing you to accurately determine the distance to objects in the scene. If, after entering the calculated parameters and subjecting the uploaded images to analysis, the final distance agrees (30 or 50 cm), then the measurement has been made correctly. This was the basic calibration of the cameras using images from two distances, which provided the necessary camera parameters, namely the focal length and angle tangent, which will be used to estimate the distance of objects in other positions as well. In short, the calibration of the system consisted of taking images from both cameras at two known distances from the object. On the basis of the known parameters of the cameras, the distances between them, and image analysis, the necessary system parameters were determined. With these, the system was able to calculate the distances to objects based on new images at other distances. The calibration process is shown in the flow chart in Figure 2.

The final element is the calculation of the distance of the object from the cameras. This was calculated based on Equation (1):

$$d = f_l + \frac{d_c \cdot N}{2 \cdot p \cdot tan\alpha} \quad (1)$$

where:

d—distance of the object from the cameras [cm];
f_l—focal length of the cameras [cm];
d_c—distance between cameras [cm];
N—width of the image [px];
p—difference in the position of corresponding points in the images from two cameras (parallax in stereoscopy) [px];

tanα—tangent of the angle of view of the camera.

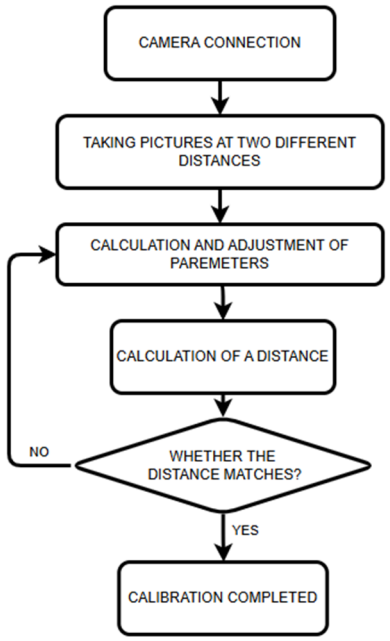

Figure 2. Flow chart of the calibration process.

This formula is based on the principles of trigonometry and the geometry of stereoscopic vision. The tangent of the angle helps to account for the divergence of the optical axes of the two cameras, and the parallax (p) helps determine the precise difference in position of the same object in the two images. The focal length (f_l) and the image width (N) are essential for scaling distances and accurately placing an object in three-dimensional space. To better understand the equation, a diagram was created, as shown in Figure 3.

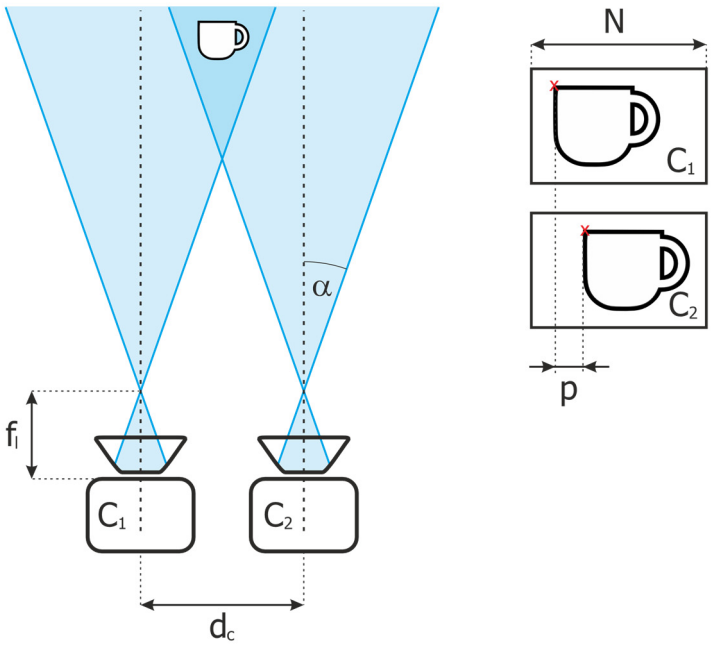

Figure 3. Illustration of the calculation method.

The methodology of the study included a test conducted on two objects: a white teacup and a LEGO human figure, as the previously mentioned identifying models detect

them without any problem. The cup is 67 mm high and 105.5 mm wide (measured with the handle), with a diameter of 86.4 mm. The LEGO figure, on the other hand, is 40.3 mm high, 24.5 mm long, and 7.7 mm wide. Thanks to the discrepancy in the dimensions of the two test pieces, it will also be possible to assess the effect of the objects' dimensions on the precision of the measurements. The experiment was carried out for distances ranging from 20 to 70 cm, increasing the distance by 2.5 cm. Three measurements were made for each distance, allowing a large amount of data to be collected. Such accuracy and repeatability of the measurements enable reliable analysis and verification of the results, assessing the estimation of the distance between cameras and objects. With such a wide range of data, it is also possible to determine the impact of different distances on the accuracy of the measurement and the calibration of the system.

The 20–70 cm distance range was chosen for its potential application in medicine, particularly in the analysis of mandibular mobility, where cameras mounted at a distance of about 30 cm can accurately monitor the movements of this structure. The selected range is also optimal for precise measurements in scenarios where objects are in close range, such as the analysis of limb movements or observation of small parts in rehabilitation. In addition, this range is ideal for monitoring isolated systems, such as finger or hand movements, where high accuracy is required close to the measuring device. In this way, the system can be used for both the detailed observation of small objects and the study of precise movements in confined spaces, expanding its potential applications in physiotherapy and rehabilitation.

3. Results

The results are presented below, which were generated to verify the accuracy of object distance calculations in the stereovision system. The test was performed on two objects: a white cup and a LEGO human figure. Measurements were made at distances ranging from 20 to 70 cm, taking three measurements every 2.5 cm for each distance. This resulted in a large amount of data, which were analyzed in detail. The absolute error, or the difference between the actual distance and the average measurement value, and the relative error, or the absolute error expressed as a percentage of the actual distance, were calculated. Example measurement results are shown in Figures 4 and 5. The results obtained are presented below in Tables 1 and 2 and Figure 6, which will be used to assess the precision of the system, as well as to identify potential sources of measurement error.

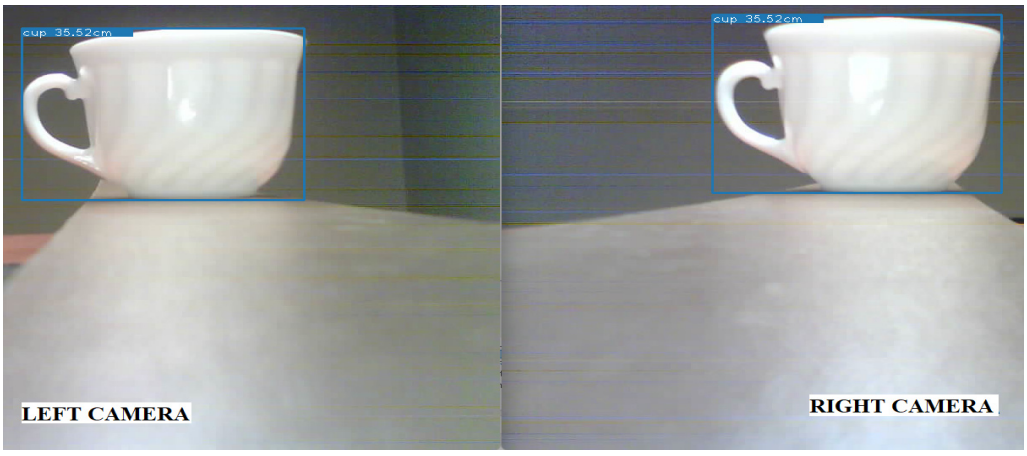

Figure 4. Measurement of the distance of a cup sample.

When analyzing the measurement results for the cup, it can be seen that the absolute errors are relatively low in most cases, especially for a certain range of distances. The largest absolute error is 4.43 cm at a distance of 70 cm, which translates into a relative error of 6.32%. The smallest absolute error occurs at a distance of 50 cm and is only 0.02 cm, representing a relative error of 0.05%. For the LEGO man, the absolute errors are larger

than for the cup, especially at longer distances. The largest absolute error is 4.81 cm at a distance of 70 cm, which translates into a relative error of 6.88%. In contrast, the smallest absolute error occurs at a distance of 45 cm and is 0.23 cm, representing a relative error of 0.52%.

Analyzing the measurement data for the cup, we note that there is a certain range of distances where measurement errors are much lower than in other cases. This range is from 25 to 60 cm, where the average absolute error for this range is 0.4 cm and the relative error is 0.95%. In this range, the individual absolute errors are consistently less than 1 cm, indicating the high precision of the measurement system in this distance range, which translates into relative errors of mostly less than 1%, with the highest in this range being only 1.61% at a distance of 45 cm and the lowest being 0.05% at a distance of 50 cm. We can see that in the vast majority of measurements in this range, the absolute error is less than 0.5 cm, which is an excellent result. In particular, measurements in the range from 27.5 to 45 cm have extremely low errors.

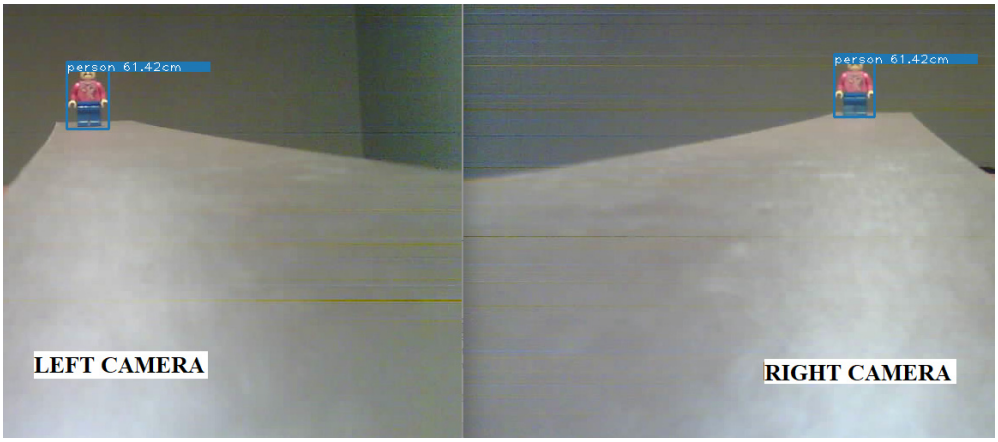

Figure 5. Measurement of the sample distance of a LEGO human figure.

Table 1. Results of cup distance measurements using a stereo camera.

Actual Distance [cm]	Measurement 1 [cm]	Measurement 2 [cm]	Measurement 3 [cm]	Average Measur. [cm]	Absolute Error [cm]	Relative Error [%]
20.0	17.48	17.52	17.17	17.39	2.61	13.05
22.5	19.74	20.4	21.53	20.56	1.94	8.64
25.0	24.54	24.57	24.58	24.56	0.44	1.75
27.5	27.72	27.73	27.43	27.63	0.13	0.46
30.0	29.59	29.59	29.66	29.61	0.39	1.29
32.5	32.27	33.16	32.75	32.73	0.23	0.70
35.0	35.52	34.84	35.59	35.32	0.32	0.90
37.5	38.18	37.87	37.72	37.92	0.42	1.13
40.0	40.42	40.43	40.27	40.37	0.37	0.93
42.5	43.06	42.06	42.87	42.66	0.16	0.38
45.0	45.71	45.84	45.62	45.72	0.72	1.61
47.5	48.48	47.67	47.56	47.90	0.40	0.85
50.0	50.25	50	49.82	50.02	0.02	0.05
52.5	53.18	52.9	53.23	53.10	0.60	1.15
55.0	54.75	54.72	54.83	54.77	0.23	0.42
57.5	56.69	56.23	57.21	56.71	0.79	1.37
60.0	59.27	59.45	59.00	59.24	0.76	1.27
62.5	60.37	60.68	60.23	60.43	2.07	3.32
65.0	63.54	62.85	64.20	63.53	1.47	2.26
67.5	65.54	65.49	65.38	65.47	2.03	3.01
70.0	65.35	65.63	65.74	65.57	4.43	6.32

Table 2. Results of LEGO human figure distance measurements using a stereocamera.

Actual Distance [cm]	Measurement 1 [cm]	Measurement 2 [cm]	Measurement 3 [cm]	Average Measur. [cm]	Absolute Error [cm]	Relative Error [%]
20.0	17.81	17.66	17.51	17.66	2.34	11.70
22.5	20.90	20.75	20.80	20.82	1.68	7.48
25.0	23.29	23.3	23.44	23.34	1.66	6.63
27.5	26.29	26.26	26.26	26.27	1.23	4.47
30.0	29.31	28.85	29.99	29.38	0.62	2.06
32.5	32.12	32.16	32.15	32.14	0.36	1.10
35.0	34.60	34.8	34.54	34.65	0.35	1.01
37.5	36.63	37.14	36.83	36.87	0.63	1.69
40.0	38.95	39.12	39.41	39.16	0.84	2.10
42.5	41.77	41.93	41.03	41.58	0.92	2.17
45.0	44.47	44.99	44.84	44.77	0.23	0.52
47.5	47.00	47.02	47.01	47.01	0.49	1.03
50.0	48.17	48.03	47.93	48.04	1.96	3.91
52.5	50.93	50.59	50.64	50.72	1.78	3.39
55.0	52.62	53.15	52.98	52.92	2.08	3.79
57.5	55.46	54.84	55.68	55.33	2.17	3.78
60.0	57.09	56.93	57.33	57.12	2.88	4.81
62.5	58.83	59.26	58.7	58.93	3.57	5.71
65.0	61.75	61.21	61.42	61.46	3.54	5.45
67.5	63.87	63.89	62.99	63.58	3.92	5.80
70.0	64.87	65.02	65.67	65.19	4.81	6.88

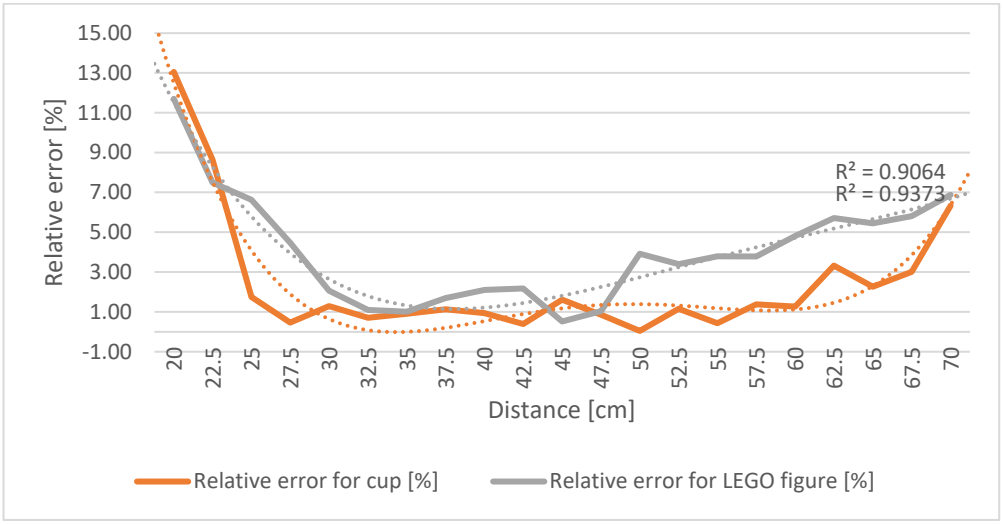

Figure 6. Comparison of the results for both objects.

The same situation is the case with the measurements of the LEGO figure; a certain range of distances is visible for which the absolute errors are much lower than 1 cm. This is the range of 30 to 47.5 cm, and the average absolute error of this separated interval is 0.56 cm, which translates into a relative error of 1.46%. The measurements obtained indicate high measurement precision, but only one measurement has a relative error of less than 1%, and this is exactly 0.52% at a distance of 45 cm, where the absolute error is 0.23 cm. In this separated range, the highest errors are obtained at a distance of 42.5 cm, where the absolute error is 0.92 cm and the relative error is 2.17%, which is still a good result.

Taking into account the entire measurement range, the average absolute error for the cup is 0.98 cm, and the relative error is 2.42%. These results indicate that the measurement system for the cup, which is a larger object, has fairly high precision. The average absolute error suggests that the measurement readings are usually close to the actual values, with a deviation of less than 1 cm. The relative error of 2.42% means that most measurements

differ from the actual values by just over 2%, which can be considered an acceptable level of precision for many applications.

For the LEGO figure, which is a smaller object, the absolute and relative errors are higher than for the cup. The average absolute error of 1.81 cm and the relative error of 4.07% are significantly higher, indicating that the system is less accurate for smaller objects. Such a result may suggest that the system has difficulty accurately detecting and measuring smaller objects, which may be related to the lower precision of segmentation or recognition of details in images.

The limitations of the methods used are due to several factors. First, the quality of the stereo camera calibration is crucial to the precision of the measurements, and inaccuracies in calibration can lead to errors in depth estimation, which directly affects the accuracy of distance measurements. A practical solution to this problem could be the use of more advanced calibration techniques or the use of cameras with built-in calibration support. Second, the Mask R-CNN algorithm itself, while effective in segmentation and object recognition, does not always accurately reflect the shape and boundaries of objects, especially those that are small, poorly visible, or when the quality of the cameras is not high enough. More advanced segmentation algorithms or streamlined versions of Mask R-CNN, tailored to detect small objects, could be considered to help reduce these inaccuracies. An additional limitation is the selection of a low-cost camera module, such as the OV2640, which also limits precision. Switching to a higher-resolution camera module could significantly improve accuracy, especially for objects with more complex shapes and contours.

Despite these limitations, the study confirms that stereo cameras can be used to measure distances and meet the work's objectives of determining the accuracy of the measurement system and identifying the range of distances where measurements are most precise. Future research should focus on improving the system calibration process and testing different segmentation algorithms tailored to objects of different sizes and shapes, potentially using the latest computer vision models. Testing different higher-quality cameras could provide valuable data on the impact of resolution on measurement accuracy. Additionally, exploring sensor fusion techniques that combine data from different types of sensors could improve the accuracy of object and depth detection under varying conditions. These directions could significantly expand the capabilities and precision of stereoscopy-based measurement systems in a variety of applications.

It is also advisable to conduct further research on the effects of various factors, such as the lighting, shape, and texture of objects, on the accuracy of measurements. The consideration of different stereo camera models and alternative measurement methods could lead to the further optimization of the system and expand its usefulness in various fields.

4. Potential Application of the System

Distance measurement systems play a key role in medicine, especially in the context of evolving diagnostic, surgical, and rehabilitation technologies. Distance measurement systems, such as stereo cameras or 3D sensors, are used to monitor patient movements, analyze posture, and precisely guide surgical instruments in minimally invasive procedures. Distance measurement with stereo cameras is also finding significant use in analyzing mandibular movement, which is particularly important in dental, orthodontic, and oral surgery diagnostics. Stereo cameras can accurately track and record three-dimensional images, allowing for the precise monitoring of dynamic jaw movements. This is important to examine temporomandibular joint dysfunction, plan orthodontic treatment, or evaluate progress after surgery.

The system's current capabilities, with its precision in the distance ranges of 25–60 cm for larger objects and 30–47.5 cm for smaller objects, indicate its usefulness in monitoring larger areas of a patient's body or objects such as prostheses or rehabilitation aids. However, for more precise medical applications, especially in internal diagnostics or microscopic surgical interventions, further optimization of the algorithm and the calibration system is needed. The use of more advanced error correction algorithms, better camera calibration,

or the use of more precise cameras could significantly reduce measurement errors for smaller objects, increasing the potential application of the system in precision medical interventions. Implementing more precise cameras, capable of operating from just a few millimeters, could significantly increase the system's range of applications. Such cameras would allow for the more accurate monitoring of fine anatomical structures and support complex medical procedures, where measurement at the microscopic level counts. With appropriate adjustments, the system could not only monitor patient movements, but also support physicians during complex surgical procedures that require extremely accurate distance measurement and instrument positioning. With appropriate adjustments, including calibration and the improvement of detection algorithms, the system could also become a useful diagnostic tool in dentistry and oral surgery, offering a noninvasive method for analyzing mandibular biomechanics and assisting in the treatment of patients suffering from masticatory organ problems.

Stereo camera-based distance measurement technology, such as the one described in this paper, has the potential to revolutionize many aspects of medicine, including endoscopy. The addition of an additional camera to existing endoscopic systems would make it possible to obtain precise information about the distance of objects in the field of view, which would significantly improve control and precision during procedures. Such a system, enhanced by stereoscopic analysis, could provide surgeons not only with images, but also with accurate data on the position of surgical instruments relative to the patient's tissues. Such a solution would support the entire surgical process, allowing for the more precise manipulation of instruments in tight and hard-to-reach spaces, which is important in many endoscopic procedures. The introduction of such technology could also reduce the risk of damaging sensitive anatomical structures, as surgeons would have access to an additional layer of information to help assess the depth and spatial arrangement of tissues. With the ability to precisely determine the distance between the tools and anatomical structures, the surgeon could make better rational decisions during surgery, which could ultimately reduce the duration of procedures, reduce the risk of complications, and speed up patient recovery. Integrating stereoscopic technology with endoscopic systems opens up new possibilities for precision surgery and diagnostics, supporting the surgeon not only in visual orientation, but also in accurate positioning of tools, which can significantly improve the safety and efficiency of medical procedures.

5. Conclusions

This paper presents the methodology and measurement stand created to test the distance of objects from a stereo camera. By comparing absolute and relative errors, the accuracy of the system was evaluated and the distance ranges where the measurements are most precise were identified. This was evaluated on the basis of the results obtained from measurements in the range of 20 to 70 cm, where three measurements were taken every 2.5 cm.

The conclusions of the study of the precision of a measurement system based on stereo camera technology indicate significant differences in measurement accuracy depending on object size and distance from the camera. Analyzing the results for two different objects—a cup and a LEGO man figurine—it was noted that the system had different levels of precision depending on the object size and distance range.

For measurements of the cup, which is a larger object, the system showed very high precision in the distance range of 25 to 60 cm. In this range, absolute errors were generally less than 1 cm, which translated into an average absolute error of 0.4 cm and a relative error of 0.95%. The highest precision, characterized by a minimum absolute error of 0.02 cm (relative error of 0.05%), was achieved at a distance of 50 cm. It is important to note that this is one of the distances at which the cameras were calibrated. It should be noted that an absolute error of less than 0.5 cm occurred in most of the measurements in this range, demonstrating the exceptional precision of the system under these conditions. The largest absolute errors were recorded at a distance of 70 cm, where they were 4.43 cm (relative

error of 6.32%). This showed that the accuracy of the system decreased at longer distances, which may be due to technological limitations or the specifics of the object itself and its reflective properties.

The LEGO figurine, which is a smaller object, had overall larger measurement errors compared to the cup. The largest absolute error was 4.81 cm at a distance of 70 cm, representing a relative error of 6.88%. The lowest absolute error was recorded at a distance of 45 cm and was 0.23 cm (relative error of 0.52%). The separated range of distances from 30 to 47.5 cm had an average absolute error of 0.56 cm and a relative error of 1.46%, suggesting that measurements in this range were more precise compared to measurements at other distances. The highest error in this range was 0.92 cm at a distance of 42.5 cm, which translates into a relative error of 2.17%. These results showed that the system tended to produce better results at closer distances for smaller objects.

By comparing the two cases, it can be concluded that the measurement system is more precise for larger objects, which is evident in the lower average absolute and relative errors. The average absolute error for the cup was 0.98 cm and the relative error was 2.42%, which is an acceptable level of precision for many applications. For the LEGO man figurine, the average absolute error was 1.81 cm, and the relative error was 4.07%, suggesting a greater variation in measurement. These results point to the need for further optimization of the system, especially for measuring smaller objects at longer distances.

In conclusion, the study has provided valuable data on the capabilities and limitations of the stereo camera-based measurement system. The system's high precision over certain distance ranges and for larger objects shows its potential for precision measurement applications, but higher errors for smaller objects indicate areas that require further research and improvement. These results can serve as a basis for the further optimization of the technology and the adaptation of the system to different applications, which is key to its versatile use in practice.

This article focuses on the use of only one neural network, Mask R-CNN, for object detection, allowing a thorough analysis of its capabilities and effectiveness. Nevertheless, the authors are not limited to this method exclusively. New studies using other neural networks are planned for the future. Their results will be used to write the second part of this article, where it will be possible to compare the methods used and evaluate their effectiveness in a broader context.

In future research to improve the measurement system, it is worth considering more advanced cameras, such as the ArduCam B0386 models dedicated for Raspberry Pi, which offer a higher resolution and greater range and depth precision, which can significantly affect the accuracy of measurements. In addition, to improve the quality of segmentation and object recognition, it would be worth testing more advanced algorithms such as U-Net or DeepLab, which work well for the precise segmentation of objects with complex shapes and may be better suited to detecting smaller structures. An important aspect of future research would also be to test the system under a variety of lighting conditions to assess the robustness of the measurements to varying light levels and to eliminate errors associated with low contrast in harsh conditions.

Author Contributions: Conceptualization, S.N. and A.Z.; methodology, S.N.; software, A.Z.; validation, S.N. and A.Z.; formal analysis, S.N.; investigation, A.Z.; resources, A.Z.; data curation, A.Z.; writing—original draft preparation, A.Z.; writing—review and editing, S.N.; visualization, A.Z.; supervision, S.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

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

References

1. Dawood, Y.; Ku-Mahamud, K.; Kamioka, E. Distance measurement for self-driving cars using stereo camera. In *International Conference on Computing and Informatics*; Universiti Utara Malaysia: Kuala Lumpur, Malaysia, 2017; Volume 1.
2. Zaarane, A.; Slimani, I.; Al Okaishi, W.; Atouf, I.; Hamdoun, A. Distance measurement system for autonomous vehicles using stereo camera. *Array* **2020**, *5*, 100016. [CrossRef]
3. Jokela, M.; Kutila, M.; Le, L. Road condition monitoring system based on a stereo camera. In Proceedings of the 2009 IEEE 5th International Conference on Intelligent Computer Communication and Processing, Cluj-Napoca, Romania, 27–29 August 2009; IEEE: Piscataway, NJ, USA, 2009; pp. 423–428. [CrossRef]
4. Kusuworo, A.; Widodo, C.E. Distance measurement with a stereo camera. *Int. J. Innov. Res. Adv. Eng.* **2017**, *4*, 24–27.
5. Li, Z.; Vu, A.N.; To, J.K. *The Stereo Camera-A Cost-Efficient User Controlled Medical Instrument for Ophthalmic Use*; Whitney High School: Cerritos, CA, USA, 2021.
6. Hernandez, A.; Basset, O.; Bremond, A.; Magnin, I.E. Stereoscopic visualization of three-dimensional ultrasonic data applied to breast tumours. *Eur. J. Ultrasound* **1998**, *8*, 51–65. [CrossRef] [PubMed]
7. Girshick, R. Fast R-CNN. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 1440–1448. [CrossRef]
8. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 1137–1149. [CrossRef] [PubMed]
9. He, K.; Gkioxari, G.; Dollar, P.; Girshick, R. Mask R-CNN. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 2980–2988. [CrossRef]
10. Yi, Z.; Lian, J.; Liu, Q.; Zhu, H.; Liang, D.; Liu, J. Learning rules in spiking neural networks: A survey. *Neurocomputing* **2023**, *531*, 163–179. [CrossRef]
11. Wu, X.; He, W.; Yao, M.; Zhang, Z.; Wang, Y.; Xu, B.; Li, G. Event-based Depth Prediction with Deep Spiking Neural Network. *IEEE Trans. Cogn. Dev. Syst.* **2024**, *16*, 2008–2018. [CrossRef]
12. Mehendale, N. Object Detection using ESP 32 CAM. *SSRN Electron. J.* **2022**. [CrossRef]
13. Verma, K.; Charan, G.S.; Pande, A.; Abdalla, Y.A.; Marshiana, D.; Choubey, C.K. Internet Regulated ESP32 Cam Robot. In Proceedings of the 2023 7th International Conference On Computing, Communication, Control And Automation (ICCUBEA), Pune, India, 18–19 August 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1–5. [CrossRef]
14. Lin, T.-Y.; Dollar, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature Pyramid Networks for Object Detection. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 936–944. [CrossRef]
15. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 3431–3440. [CrossRef]
16. Mao, W.; Yu, H. Object Recognition and Location Based on Mask R-CNN and Structured Light Camera. In Proceedings of the 2019 International Conference on Computer, Network, Communication and Information Systems (CNCI 2019), Qingdao, China, 27–29 March 2019; Atlantis Press: Paris, France, 2019. [CrossRef]