Object distance analysis using convolutional neural network (CNN) method based on stereo vision

B. Hilda Nida Alistiqlal[1], I Wayan Sudiarta[1*], Susi Rahayu[1]

[1]Department of Physics, Faculty Mathematics and Natural Sciences, University of Mataram, Mataram, NTB 83125, Indonesia

wayan.sudiarta@unram.ac.id

Received 18 April 2024, Revised 29 July 2024, Published 30 September 2024

Abstract: Blind or semi-blind people require additional devices or assistants such as a cane or a guide dog to navigate in their daily activities. In robotic or automation technology such as self-driving cars also require knowing objects in front of them accurately. The purpose of this paper is to study the use of stereo cameras as a distance measuring device and to determine the optimal camera configuration and its accuracy. The method used to determine distance is the convolutional neural network (CNN) method based on stereo vision. For training and validation data, we recorded images of a ball in random positions in front of the stereo cameras. This test uses two steps, the first is to recognise the camera object using the CNN method, then the second is to measure the stereovision-based object with an input image of 1350 image pairs. With 160 images used as CNN training data and the remaining 40 images as CNN validation data, the entire dataset is also taken to be used as training data for ball distance. Based on the results of the study, it was found that the model successfully predicted the image with 97% accuracy. The accuracy of the optimal distance measurement results at a distance between cameras of 11 cm is found to be 68.03% within a distance of 1 m, for a distance between cameras of 15 cm is found to be 94.63% within a distance of 2 m, and for a camera distance of 19 cm is found to be 99.61% for a distance of 3 m objects.

Keywords: Convolutional Neural Network (CNN); Measuring Distance; Stereo cameras.

1. Introduction

The world is rapidly developing in the field of robotic automation, which greatly facilitates human work. Robots are increasingly being made to imitate humans in order to simplify and replace human activities such as walking, hearing, seeing, talking, and so on. One example is the concept of self-driving car technology, in which the car is operated by a specially designed system that can act like a driver by moving the steering wheel, brakes, gas, and so on. This can reduce the number of traffic accidents on the highway. Data shows that in 2020, there were 100,028 traffic accident cases in Indonesia alone. Therefore, technological developments in the development of smart cameras for road vehicles, especially cars, are expected to be a solution in reducing traffic accident cases. In order to do so, robots or cars must be able to avoid obstacles such as objects or holes.

The camera is an important component in this process. The camera is used to take pictures of the area around the object. Various studies have been conducted to explore good techniques in the development of robots that measure distances, such as using ultrasonic sensors (Arsada, 2017), infrared sensors (Ramadhan, 2012), and lasers (Sabirin et al., 2020).

Basically, stereo vision, a development of computer vision, aims to imitate the way humans see or human vision. Human vision is actually very complex. It starts with humans seeing objects with their sense of sight, and then the image of the objects is transmitted to the brain, where the objects are interpreted. This interpretation can then be used to make decisions. (Kusumanto & Tompunu, 2011)

There are several applications related to computer vision, including image classification, image recognition, and object detection. (Dewi, 2018) Object detection is still under development, but it is progressing rapidly in its goal of imitating humans' ability to see and understand objects. One of the methods used in object detection is the Convolutional Neural Network (CNN).

Convolutional Neural Network (CNN) is a type of deep learning because of its layered structure. Deep learning is a branch of machine learning that can teach computers to perform tasks like humans. Just like humans, computers can learn from a process called training. CNN is a deep learning method that can automate feature extraction. This method is widely used in image extraction, where the extracted information is then processed by the computer, including image classification with a high level of accuracy. Therefore, in its development, a CNN model can be created to detect objects (Valueva et al., 2020)

2. Experimental Method

This study aims to detect objects and object distances using two cameras. Data containing images of a ball with actual measured distances were collected for use in the study as primary data. This study uses data captured by two cameras with a resolution of 1280x480 pixels, which is then used as an image dataset and ball distance to be analyzed. The initial data that has been collected is then used as two datasets: the first dataset is the actual pixel image data. The results of the two datasets will be processed to determine which part of each pixel has a spherical pattern and its location in the pixels in the image that can capture images in real-time to determine whether the image captured by the camera is a ball or not using the CNN model with the YOLOv5 algorithm. All datasets will be trained to recognize and learn the pattern of ball location and ball distance from the camera using the pixel difference between the bounding box point of the right camera and the left camera. The camera design scheme can be seen in Figure 1, and the results of the camera's vision can be seen in Figure 2.

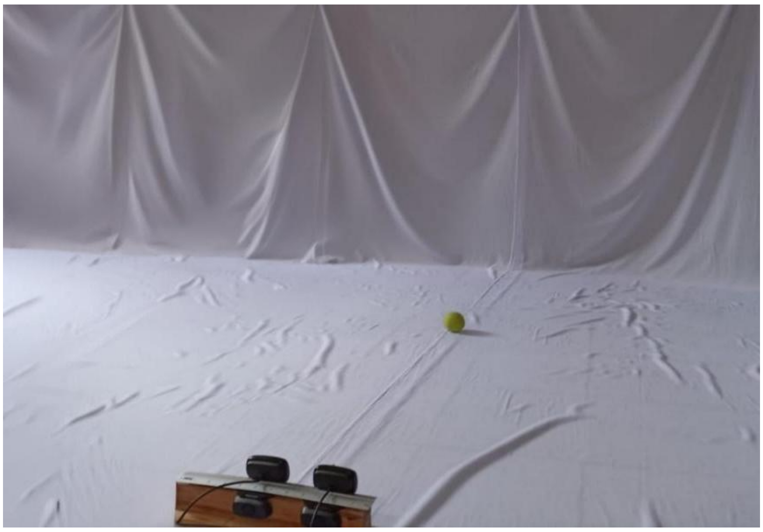

Figure 1. Configuration of a stereo camera with a tennis ball for collecting training and validation images.

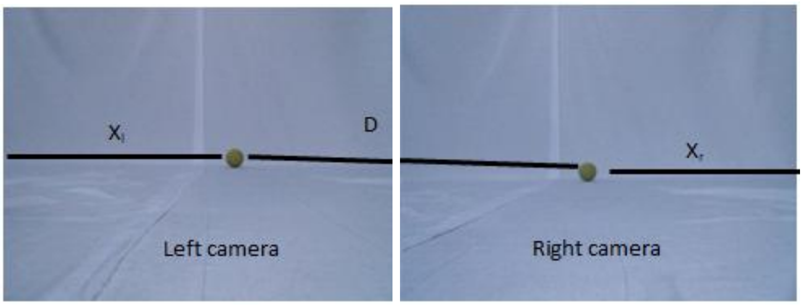

Figure 2. A pair of images captured by the two cameras

This section is a triangulation that can be modelled as shown in Figure 3 below.

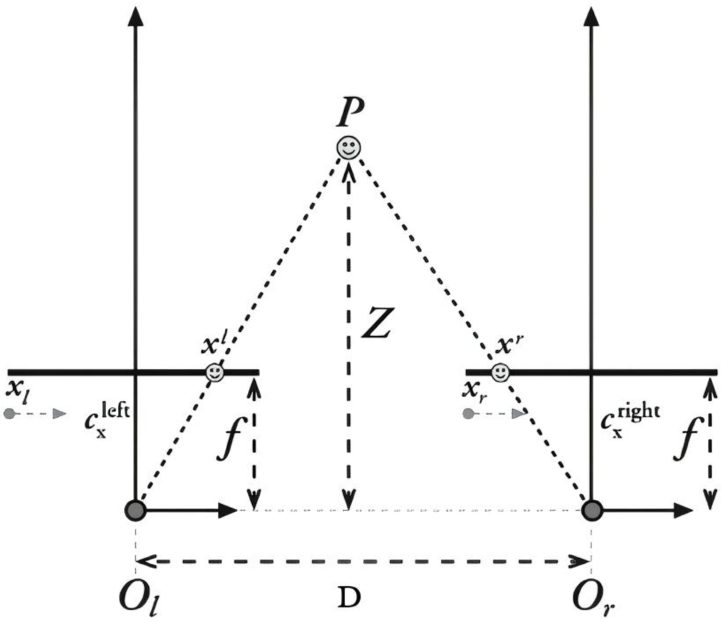

Figure 3. Stereo Vision Triangulation Modeling (Adhiem et al., 2021)

Without tilting the camera, the ball is placed in front of the camera at a predetermined distance. The camera captures the ball by processing the model created for the stereo vision programming and generates the distance between the camera and the moving object using the K-NN algorithm and the Chebhysev method. The Chebhysev method calculates the pixel difference by taking only the x-axis of the object image captured by the two cameras in real-time.

3. Result and Discussion

The results are discussed in the form of images, graphs, and values from the calculation of the error rate and accuracy of the actual object distance with the prediction of the object distance that has been made. As well as the respective distance between the camera and the image pixels.

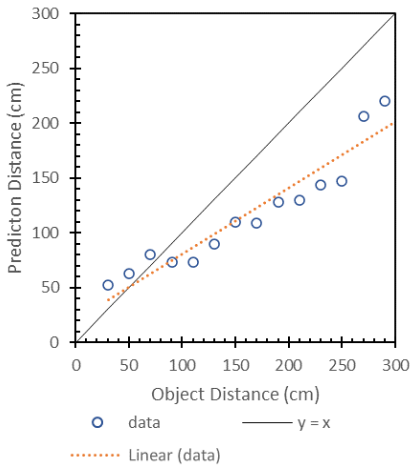

Figure 4. Measurement results for d = 11 cm

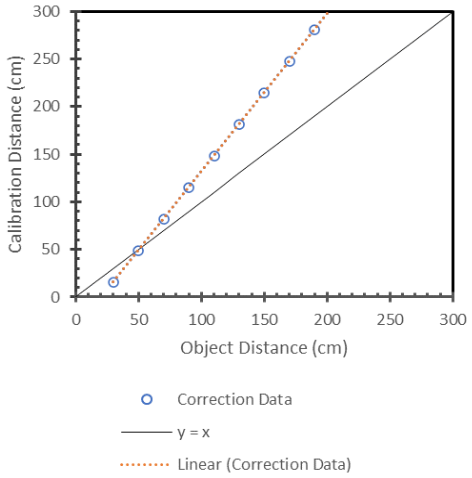

Figure 5. Measurement calibration results for d = 11 cm

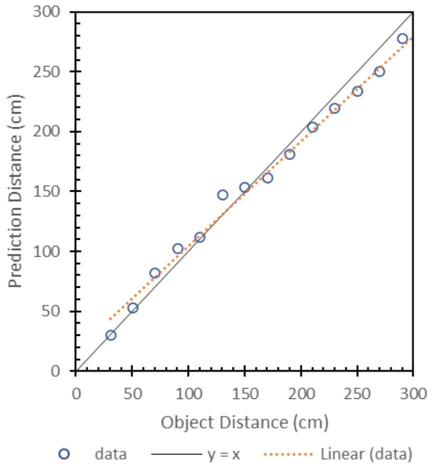

Figure 6. Measurement results for d = 15 cm

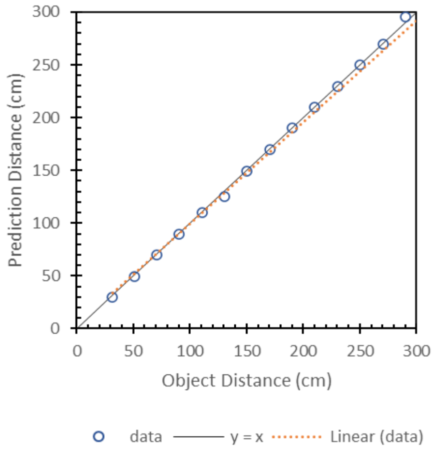

Figure 7. Measurement results for d = 19 cm

In Figure 4, it can be seen that the predicted object distances do not match the actual object distance which can be seen from the data that is not on the y = x-axis. However, the results of the prediction of object distances have shown a linear line pattern, except at distances above 150 cm. Instability in readings can occur as a result of poor camera setup and unstable data retrieval from the camera. This can be corrected by calibrating the distance prediction results using the correction function which can be obtained from the linear function regression curve to obtain the results that can be seen in Figure 5.

In Figure 6, for d = 15 cm it is ideal for measurements up to 200 cm and is better at taking measurements at an object distance of 100 cm compared to d = 11 cm. It can be seen that the predicted object distance does not match the actual object distance, which can be seen from the data that is not on the y = x-axis. However, the results of the prediction of object distances have shown a linear line pattern. The existence of an inaccurate value at d = 15 cm can be caused by the results of a bad camera setup and the instability of data retrieval from the camera when conducting trials in real-time.

In Figure 7, for d = 19 cm, it can detect the distance of the object measurement to the camera from a distance of 30 cm to 300 cm ideally and is better than d = 15 cm, also close to the original distance and is on the y = x-axis. Data at d = 19 cm has shown a linear line pattern. After testing for each distance between cameras ranging from 30 cm to 310 cm, it was found that the best results were at d = 19 cm for the actual image type, so researchers conducted trials to determine the optimal value that can be produced by the distance between cameras. The results obtained can be seen in Figure 8 as follows.

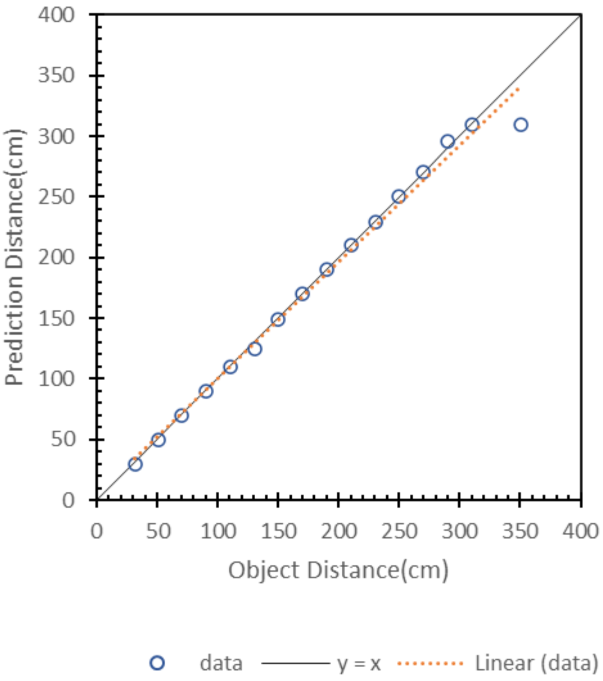

Figure 8. The optimal distance for d = 19 cm for the actual type of image

It was found that the optimal distance that can be measured by each value of d has a different distance. The optimal distance itself is the predicted distance produced by two cameras with a good level of accuracy and the smallest error. To see more clearly the average accuracy of each image with different camera distances, results can be seen in Figure 9.

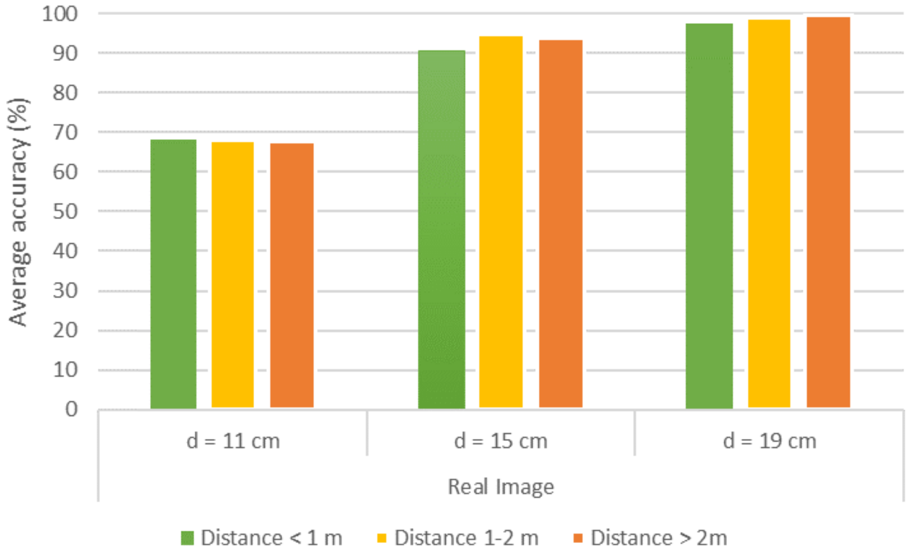

Figure 9. Average accuracy of the stereo camera for three values of camera distances.

4. Conclusion

Based on the results of the research and discussion, it can be concluded that the distance between the cameras affects the results of measuring object distances based on stereo vision. The greater the distance between the cameras and the state of objects that can be captured at the viewing distance of the two cameras, the better the measurement results. In addition, the camera resolution affects the results of measuring object distances. The higher the image resolution, the more accurate the measurement results. The best maximum distance accuracy results were obtained at a distance between cameras of 19 cm of 99.61% for an object distance of 3 m.

Acknowledgements

We thank Ramadian Ridho Illahi who helped correct Python codes for data processing.

References

Arsada, B. (2017). Aplikasi Sensor Ultrasonik Untuk Deteksi Posisi Jarak Pada Ruang Menggunakan Arduino Uno. Jurnal Teknik Elektro, 6(2), 1–8.

Ramadhan, R. (2012). Sistem Pendeteksi Objek untuk Keamanan Rumah dengan Menggunakan Sensor Infra Red. 1–17.

Sabirin, M. S., Magdalena, R., & Susatio, E. (2020). Prototipe Pendeteksi Jarak Menggunakan Kamera Smartphone Dan Laser (Prototype Distance Detection Using a Smartphone Camera and Laser With Hsv Color Model). 7(2), 4127–4133.

Kusumanto, R. D., & Tompunu, A. N. (2011). Pengolahan Citra Digital Untuk

Mendeteksi Obyek Menggunakan Pengolahan Model Normalisasi RGB. Semantik, 1(1), 37–72.

Dewi, R. S. (2018). Deep Learning Object Detection Pada Video. Deep Learning Object Detection Pada Video Menggunakan Tensorflow Dan Convolutional Neural Network.

Valueva, M. V., Nagornov, N. N., Lyakhov, P. A., Valuev, G. V., & Chervyakov, N. I. (2020). Application of the residue number system to reduce hardware costs of the convolutional neural network implementation. Mathematics and Computers in Simulation, 177, 232–243. https://doi.org/10.1016/j.matcom.2020.04.031

Adhiem, M. A. N., Nazaruddin, Y. Y., & Zahra, N. (2021). Pengembangan Trem Otonom Tanpa Rel. Jurnal Otomasi Kontrol Dan Instrumentasi, 13(2), 125–133. https://doi.org/10.5614/joki.2021.13.2.8