

Obstacle Avoidance Method using Stereo Camera for Autonomous Robot

Nabeel Kahlil Maulana¹, Widodo Budiharto^{2*}, Hanis Amalia Saputri³

¹⁻³Computer Science Department, School of Computer Science,
Bina Nusantara University,
Jakarta, Indonesia 11480
nabeel.maulana@binus.ac.id; wbudiharto@binus.edu;
hanis.saputri@binus.ac.id

Abstract—This paper presents the development and implementation of an obstacle avoidance system for an autonomous robot using a stereo camera setup. The system enables the robot to navigate its environment safely by identifying obstacles and making real-time movement decisions based on depth perception. The stereo vision configuration allows the robot to estimate distances through disparity computation and polynomial linear regression modeling. The proposed algorithm performs stereo matching, image rectification, and depth estimation to generate disparity maps representing obstacle distances. The robot uses this information to figure out if the items it sees are close, medium, or far away, and then it chooses the right move, such stopping, turning left, or turning right. The robot can find and avoid obstacles in different indoor settings, as shown by the experimental findings. The regression model employed for depth estimation attained a high degree of accuracy, evidenced by a R^2 value of 0.97 and a minimal mean absolute error, signifying robust reliability in distance prediction. The research validates that the amalgamation of stereo vision with regression-based distance estimate yields a resilient and economical method for autonomous navigation. This study advances the ongoing evolution of intelligent robotic systems that can execute autonomous decision-making with limited human oversight.

Keywords— *autonomous robot; obstacle avoidance; obstacle detection; stereo camera*

I. INTRODUCTION

Autonomous robot generally has two fundamental requirements: (1) an agent receive information from sensor and effect its behavior and (2) flexibility on switching between behaviors [1]. Recent research shows a rapid increase in the use of stereo vision for autonomous robots, particularly in the context of obstacle avoidance and autonomous navigation. This system allows robots to obtain real-time depth perception from two different images, enabling them to estimate the distance and position of objects in their surroundings with high precision [2].

Received: Oct. 27, 2025; received in revised form: Nov. 04, 2025; accepted: Nov. 04, 2025; available online: Nov. 04, 2025.

*Corresponding: wbudiharto@binus.edu

Seewald (2022) Evaluating two obstacle avoidance approaches in mobile robots using stereo cameras: the conventional stereo view algorithm method and an end-to-end disparity-sensitive machine learning-based network. The results show that although classical algorithms such as block matching are still efficient in simple environments, the neural network-based approach demonstrates significantly better performance in complex conditions with unstable lighting. These findings are important for your research, as systems developed using stereo matching can also be improved through the integration of adaptive disparity learning networks [3].

A study by Umam et al. (2023) introduced a stereo vision-based navigation system on an omnidirectional robot with Neuro-Fuzzy control. The research successfully demonstrated that combining artificial intelligence methods with depth data from stereo cameras can produce smoother and more responsive manoeuvres. Although the robot configuration used was different (omnidirectional versus four-wheeled), their approach to classifying distance and direction of movement can serve as a reference for improving the decision-making system in your research using Jetson Nano and L297N motor drivers [4].

The research by Nguyen et al. (2024) proposes an obstacle avoidance strategy based on deep reinforcement learning (DRL) that utilises stereo camera input for training autonomous decision-making models. This system learns from interactions with the environment to determine the optimal direction and speed of movement without the need for explicit programming. The DRL-based approach is relevant to your research because it demonstrates the potential for integrating stereo sensors and adaptive learning algorithms to improve the flexibility of robot behaviour when faced with previously unrecognised navigation scenarios [5].

Badrloo (2025) researched obstacle detection methods by combining stereo vision and deep learning, resulting in a system capable of processing images at high speed without sacrificing depth accuracy. This research reinforces the evidence that stereo-based obstacle detection systems equipped with deep learning models are superior in detecting small and complex objects in dynamic environments. These findings support your research's main argument that

combining stereo cameras with distance estimation algorithms based on statistical models or artificial intelligence can produce efficient, economical, and adaptive solutions for autonomous robot navigation [6].

This paper discusses obstacle avoidance system using stereo camera setup that retrieves information and dictates the maneuver. Obstacle avoidance is a problem where an entity, in this case a robot, avoids obstacles on its path to prevent any collision. We use stereo vision to perceive depth of captured image and analysis the distance of the objects. This setup allows the robot to analyze all projected objects in 2D plane image. The stereo vision setup uses two cameras facing onward. We designed a four wheeled robot, shown in figure 1, that uses two Logitech C270 HD Webcam and computes the information using Jetson Nano 2GB. The Jetson can send signals to L297N motor drivers to control robot movement. The robot is supplied with DC batteries and 5V 2A powerbank as shown in figure 1.

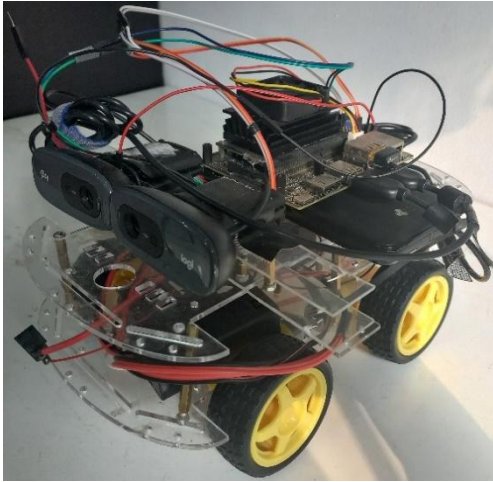


Figure 1. Four Wheeled Autonomous Robot Equipped with Stereo Camera Setup

II. PROPOSED METHOD

A. Stereo Matching

Szeliski describes stereo matching as a process of taking two or more images and estimating a 3D model of the scene by finding matching pixels in the images and converting their 2D positions into 3D depths [7]. In this case, the images are captured by two cameras that face the same scene from slightly different angle. The result of such setup will generate an image pair (consist of left image and right image). The matched features (a group of pixels that have unique patterns) of the image pair is called stereo correspondence. By knowing the correspondence feature from both images, we can perceive the depth of that feature by calculating the disparity. The disparity is taken from the difference of the position of both correspondence features.

To select which features correspond to other target features, stereo matching uses cost criterion to measure the matching likeliness. Matching cost that this paper use is *sum of absolute difference* (SAD) [8]. SAD equation is defined in (1).

$$SAD = \sum |I_l(x, y) - I_r(x + d, y)| \quad (1)$$

Where I_l and I_r denotes a local area of an image which center location is defined by horizontal x and vertical y position of the image. The size of the local area is defined by the window size. By finding the lowest SAD of all potential matching features, the value of disparity d of that feature is taken to make *disparity map* or what we call, *disparity image*. In this research, we use *stereoBM* module from OpenCV [9] to compute the disparity image.

B. Image Rectification

The process of matching stereo assumes that epipolar lines of the image pair are parallel. Rectification is a process that wrap the "raw" image pair into rectified images. The process includes tilting, scaling, and rotating of both images using fundamental matrix describes in equation (2) [10]. discusses how to find such matrix and discusses how the image rectification process is implemented using fundamental matrix [11].

$$k = \begin{pmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix} \quad (2)$$

The importance of rectification is to improve the computation time and complexity of the stereo matching algorithm. When the epipolar line of both left and right image area parallel, the number of potential features is reduced to features in horizontal line where vertical y is the same. This horizontal line is called *scanline*.

C. Depth Estimation

The value of disparity of a pixel in image describes the depth of an object. But the value is not the real-world distance. Using disparity alone to measure the distance is not sufficient. By using stereo camera triangulation, the relationship between disparity and distance is described in equation (3).

$$z = \frac{f_x b}{d} \quad (3)$$

Where f_x is horizontal focal length of the camera, baseline b is the distance of the two sensors of the stereo camera, and distance z has the inverse relation with disparity d . Therefore, the larger the value of the disparity, the closer of the object to the camera sensor. This relationship is portrayed in figure 3. The other method of finding the distance is predicting the value using a polynomial linear regression. We use this model to estimate the real-world distance by feeding it with disparity maps. It is a simple linear regression where the variables have the power degree of polynomial. The regression values can fit the data in figure X where the relationship between data is non-linear. The function of this model can be described in equation (4).

$$y(x) = \beta_0 x^0 + \beta_1 x^1 + \beta_2 x^2 + \beta_3 x^3 + \epsilon \quad (4)$$

Where β_n are coefficients, ϵ is the error term, and x^n are the variables. The value of y is dependent of the value of

independent x [12]. The other name of this model is cubic regression model or 3rd degree polynomial model (the highest power degree in x^n). In this research the model accepts disparity as x and predicts the distance as y .

C. Stereo Vision Architecture

Our stereo vision system is similar to [13]. The robot is equipped with two cameras that face frontward. This stereo vision setup allows robots to perceive depth of the objects by analyzing the matching points of captured image pair. The Figure 2 shows data flow of processing raw image pair into depth mapping.



Figure 2. Stereo Vision Data Flow

The image pair are captured with a resolution of 640x480 pixel and a color-space of grayscale. Image rectification requires few parameters to rectify the image pair using fundamental matrix. This fundamental matrix is calculated through camera calibration and stereo calibration. Both calibrations implement technique from [14] where it need few testing images that captures a chessboard patterns. the chess. After images are rectified and both epipolar lines are paralleled, the stereo matching model accepts the rectified image pair and generates disparity map. Each disparity map describes the depth of projected 2D image. Then to find the specific distance value of pixel in the disparity map, we use the trained polynomial linear regression model.

D. Proposed Obstacle Avoidance System

This paper proposed a system that consists of obstacle detection and maneuver dictation. The system classifies all objects by distance into a near obstacle, medium obstacle, and far obstacle. It uses near obstacle constant (n) and far obstacle constant (f) to determine the object type. If the object distance is below n , the system classifies the object as near obstacle and it needs to stop. Because the obstacle is too close for robots to do smooth maneuver and robots need to stop or do a special movement (rotating, backward, etc.). If the object distance is between n and f , the system classifies the object as medium obstacle and determines the turn direction to avoid collision. If the object distance is above f , it can move forward without slowing down or set the turn direction. To determine which direction to choose, this paper improves the window area technique based on [15] by increasing the number of the window. It decides the object type and finds the closest pixel location to the middle of the image. The farther the obstacle location is relative to the middle of the image, the bigger the turn direction. Equation (5) and (6) describe how much the turning direction by locating the obstacle location on image.

$$a_l = \frac{x_l}{(I_w * \frac{1}{2})} * 100\% \quad (5)$$

$$a_r = \frac{x_r}{(I_w * \frac{1}{2})} * 100\% \quad (6)$$

Disparity image is split into two sections. If an object is projected mostly on the right side, then x_l donates left object location on the image. If an object is projected mostly in the left side of the image, x_r denotes the rightest pixel object location on the image. Turn direction a_l and a_r describe how much the robot should turn to avoid collision. But if x_l and x_r overshoot the middle section, the turn direction should be some constant, e.g., 10 percent.

III. EXPERIMENTAL RESULT

Training the polynomial linear regression model requires dataset as train. The dataset consists of series of value pair disparity and distance data that can be seen in Table 1 and blue plot of figure 2. The unit value of disparity is in pixel, and the distance is in millimeter (mm). The dataset obtained by capturing the value of disparity of an object using stereo matching algorithm and pair it to the object's distance to the stereo camera. The data pairs are captured several times where each time the distance is different. In this case, we capture the data where distance from 260 mm to 2800 mm incrementally. The regression model is trained using this dataset. Table 1 shows the benchmark of the trained model against the dataset.

Table 1. Trained Model Coefficient

Coefficient	Value
β_0	0.0
β_1	-144.54135
β_2	1.7902485
β_3	-0.0071868896
ϵ	4301.4795

Table 2. Depth Estimation Model Benchmark Against Dataset

Metric	Result
<i>R-Squared (R2)</i>	0.9705985587775484
<i>Root Mean Squared Error (RMSE)</i>	138.29893228053425
<i>Mean Absolute Error (MAE)</i>	116.50050123007252
<i>Mean Absolute Percentage Error (MAPE)</i>	0.15368257320976403

Table 3. Gathered Disparity and Distance Pair Example Dataset

	Disparity (pixel)	Distance (mm)
1	142.5	260
2	128.75	300
3	107.6875	350
4	118.4375	330
5	93.375	400
10	65.00	600
15	33.5625	1200
20	14.8125	2800

To test our obstacle avoidance system, we conduct an experiment in controlled indoor environment. The environment is a room of 3m x 3m and has obstacles that can be placed temporarily. The obstacle placement is based on which maneuver to test. There are 3 scenarios to experiment which are left maneuver, right maneuver, and stop. The left and right maneuver is when the robot distance to obstacle is in within n and f and the obstacle placement is slightly off relative to robot's axis. These two scenarios are to test whether the robot can decide on which direction to choose to avoid the obstacle. The stop scenario is when the obstacle distance is below n and robot should stop. Figure 3 plots the regression result that fits into the dataset. Figure 4 shows the obstacle detection (represented by white squares) in various scenarios. Figure 5 shows the robot trajectory with maneuver scenario.

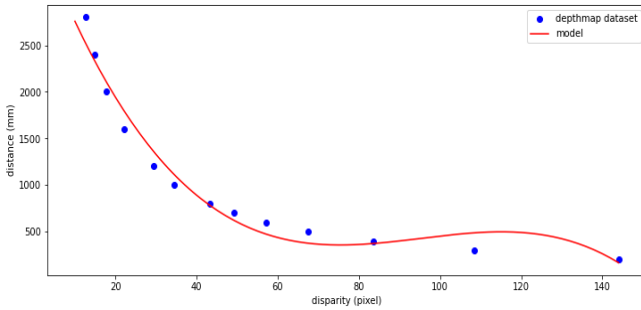


Figure 3. Depth Estimation Model Regression

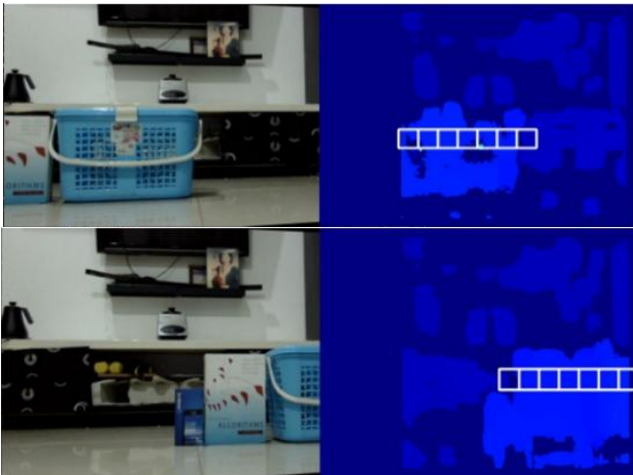


Figure 4. Obstacle Detection Sample: (Top-Left) Right Maneuver Scenario Where Obstacle is in Left, (Top-Right) Left Maneuver Scenario Where Obstacle is in Right, (Bottom) Stop Scenario Where Obstacle is too Close to Robot

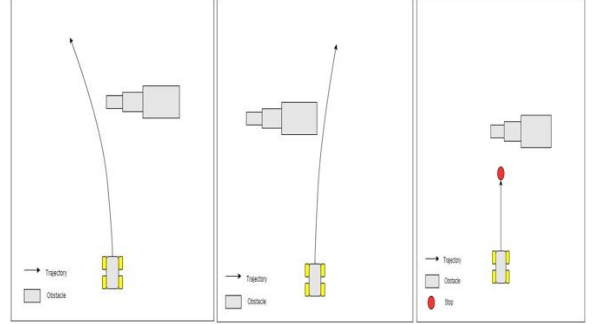


Figure 5. Robot Trajectory Testing: (Left) Left Maneuver Scenario, (Middle) Right Maneuver Scenario, (Right) Stop Scenario

IV. CONCLUSION

The autonomous robot using our proposed obstacle avoidance system can detect an obstacle and avoid any obstacle on its path. The stereo camera allows the robot to perceive the depth of an object by calculating the disparity projected pair images. Then it predicts the actual distance using polynomial linear regression model. We conduct experiments in three different scenarios to test the obstacle system.

ACKNOWLEDGMENT

This work is fully supported by BINUS University.

AUTHOR'S CONTRIBUTION

The study was conceived and designed by Widodo Budiharto. The experiments and programming are performed by Nabeel Kahlil Maulana. The writing is guided by Widodo Budiharto. The Writing - Review & Editing by Hanis Amalia Saputri. All authors read and approved the manuscript.

REFERENCES

- [1] G. Schöner, M. Dose, and C. Engels, "Dynamics of behavior: Theory and applications for autonomous robot architectures," *Robotics and Autonomous Systems*, vol. 16, no. 2, pp. 213–245, Dec. 1995, doi: 10.1016/0921-8890(95)00049-6.
- [2] H. Li, Z. Li, N. Unver, and M. R. Azimi, "StereoVoxelNet: Real-Time Obstacle Detection Based on Occupancy Voxels from a Stereo Camera Using Deep Neural Networks," *arXiv preprint arXiv:2209.08459*, 2022.
- [3] A. K. Seewald, "Evaluating Two Ways for Mobile Robot Obstacle Avoidance with Stereo Cameras: Stereo View Algorithms and End-to-End Trained Disparity-Sensitive Networks," in *Proc. 11th Int.*

- Conf. on Pattern Recognition Applications and Methods (ICPRAM), 2022, pp. 301–308.
- [4] F. Umam, M. Fuad, I. Suwarno, A. Ma'arif, and W. Caesarendra, "Obstacle Avoidance Based on Stereo Vision Navigation System for Omni-Directional Robot," *Journal of Robotics and Control (JRC)*, vol. 4, no. 3, pp. 345–354, 2023, doi: 10.18196/jrc.4317977.
 - [5] C.-H. Nguyen, Q.-A. Vu, and T. N. Pham, "Optimal Obstacle Avoidance Strategy Using Deep Reinforcement Learning Based on Stereo Camera," *MM Science Journal*, vol. 2024, pp. 6424–6433, Oct. 2024, doi: 10.17973/MMSJ.2024_10_2024063.
 - [6] S. Badrloo, "Fast and Accurate Obstacle Detection Based on Stereo Vision and Deep Learning," *ISPRS Archives – International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLVIII-G, pp. 177–184, 2025, doi: 10.5194/isprs-archives-XLVIII-G-2025-177-2025.
 - [7] R. Szeliski, *Computer Vision*. London: Springer London, 2011. doi: 10.1007/978-1-84882-935-0.
 - [8] R. A. Hamzah and H. Ibrahim, "Literature Survey on Stereo Vision Disparity Map Algorithms," *Journal of Sensors*, vol. 2016, pp. 1–23, 2016, doi: 10.1155/2016/8742920.
 - [9] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.
 - [10] Z. Zhang, "Determining the Epipolar Geometry and its Uncertainty: A Review," *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, vol. 37, no. 1, pp. 55–57, Dec. 1992, doi: 10.1177/075910639203700105.
 - [11] C. Loop and Z. Zhang, *Computing rectifying homographies for stereo vision*, vol. 1. 1999, p. 131 Vol. 1. doi: 10.1109/CVPR.1999.786928.
 - [12] J. L. Devore, *Probability and Statistics for Engineering and the Sciences*, 8th edition. Boston, MA: Cengage Learning, 2011.
 - [13] Rostam Affendi Hamzah, Rosman Abd Rahim, and Zarina Mohd Noh, "Sum of Absolute Differences algorithm in stereo correspondence problem for stereo matching in computer vision application," in *2010 3rd International Conference on Computer Science and Information Technology*, Chengdu, China, Jul. 2010, pp. 652–657. doi: 10.1109/ICCSIT.2010.5565062.
 - [14] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, no. 11, pp. 1330–1334, Nov. 2000, doi: 10.1109/34.888718.
 - [15] L. Nalpantidis, I. Kostavelis, and A. Gasteratos, "Stereo-vision-Based Algorithm for Obstacle Avoidance," in *Intelligent Robotics and Applications*, vol. 5928, M. Xie, Y. Xiong, C. Xiong, H. Liu, and Z. Hu, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 195–204. doi: 10.1007/978-3-642-10817-4_19.