Journal of Mechatronics, Electrical Power, and Vehicular Technology 16 (2025) 158-168

Journal of Mechatronics, Electrical Power, and Vehicular Technology

e-ISSN: 2088-6985
p-ISSN: 2087-3379

mev.brin.go.id

An evaluation of stereo vision for distance estimation using the SGBM algorithm in the CARLA simulator

Rizky Hamdani Sakti [a, *], Liptia Venica [a], Dewi Indriati Hadi Putri [a], Shinta Rohmatika Kosmaga [b], Estiko Rijanto [b]

[a] *Department of Mechatronics and Artificial Intelligence, Universitas Pendidikan Indonesia*
Jl. Veteran No. 8, Purwakarta, 41115, Indonesia

[b] *Research Center for Smart Mechatronics, National Research and Innovation Agency (BRIN)*
Kawasan Sains dan Teknologi (KST) Samaun Samadikun, Jalan Sangkuriang, Bandung, 40135, Indonesia

Abstract

This paper presents an evaluation of stereo vision based on the semi-global block matching (SGBM) algorithm for distance estimation in an autonomous parking scenario using the CARLA simulator. Distance-disparity regression functions are explored to enhance distance estimation accuracy. The proposed distance estimation model was evaluated using the design science research methodology (DSRM) framework, with experimental validation conducted in CARLA's promenade environment. The evaluation employed root mean square error (RMSE) and relative error metrics to assess performance. Experiments were performed within a range of 40-350 cm, which is relevant for autonomous parking applications. The experimental results show that the algorithm achieves an overall RMSE of 1.69 cm and an average relative error of 1.1 %. The findings contribute to the advancement of perception systems for autonomous vehicles, particularly in challenging environments.

Keywords: autonomous parking; stereo vision; SGBM algorithm; distance estimation; CARLA simulator; perception system.

I. Introduction

The rapid advancement of autonomous vehicle technology has intensified the demand for robust and accurate perception systems capable of operating under diverse environmental conditions. Among the various sensing modalities employed in autonomous vehicles, stereo vision systems offer significant advantages, including cost-effectiveness [1], passive operation [2], and rich environmental information extraction [3]. Distance measurement using stereo vision has become a critical component in autonomous parking systems, where precise spatial awareness is essential for safe vehicle maneuvering in confined spaces.

Traditional distance measurement techniques in automotive applications rely heavily on active sensors such as light detection and ranging (LiDAR) [4], radio detection and ranging (RADAR) [5], and ultrasonic sensors [6][7]. While these sensors provide accurate distance measurements, they are often expensive and may face limitations under certain environmental conditions [8]. Stereo vision systems, inspired by human binocular vision, offer a compelling alternative by utilizing two or more cameras to estimate depth information through triangulation principles [9].

* Corresponding Author. rizkyhamm@upi.edu (R. H. Sakti)

https://doi.org/10.55981/j.mev.2025.1284

Received 2 October 2025; revised 18 November 2025; accepted 24 November 2025; available online 22 December 2025

How to Cite: R. H. Sakti *et al.*, "An evaluation of stereo vision for distance estimation using the SGBM algorithm in the CARLA simulator," *Journal of Mechatronics, Electrical Power, and Vehicular Technology*, vol. 16, no. 2, pp. 158-168, Dec., 2025.

The accuracy of results obtained using stereo vision technology is highly dependent on the algorithm employed to process the acquired images [10]. The semi-global block matching (SGBM) algorithm represents a significant advancement in stereo vision technology, addressing many of the limitations of traditional block matching approaches. Unlike local methods that consider only small neighborhoods, SGBM incorporates global optimization techniques while maintaining computational efficiency suitable for real-time applications [11]. This capability makes SGBM particularly attractive for autonomous parking systems where both accuracy and processing speed are critical requirements.

Furthermore, to ensure the robustness of such algorithms before real-world deployment, simulation environments have become indispensable tools for development and validation. The car learning to act CARLA simulator provides a comprehensive platform for testing autonomous driving algorithms in controlled yet realistic environments [12]. CARLA's ability to simulate various weather conditions, lighting scenarios, and urban environments makes it an ideal testbed for evaluating stereo vision algorithms under diverse conditions, thereby accelerating the design and validation process for autonomous vehicle perception systems.

Several articles have reported implementation of the SGBM algorithm, including vehicle speed measurement [13], pavement pothole detection [14], and obstacle detection in autonomous vehicle navigation [15][16]. However, no study has reported its adaptation in the CARLA simulator environment. The CARLA simulator features a promenade environment that provides unique challenges for autonomous vehicle perception systems due to factors such as varying lighting conditions, reflective surfaces, and atmospheric effects. Object detection and tracking may struggle under these conditions, necessitating algorithm enhancements and thorough evaluation [17].

This research addresses the critical need for reliable distance measurement systems in autonomous parking applications by proposing an enhanced SGBM algorithm specifically optimized for promenade environments. The primary contributions of this work include: (1) the development of an improved SGBM algorithm with enhanced accuracy for distance measurement, (2) comprehensive evaluation using multiple accuracy metrics in the CARLA simulator environment, and (3) validation of the proposed system's suitability for autonomous parking perception systems in the critical distance range of 40-350 cm.

II. Materials and Methods

A. Stereo vision fundamental

A stereo vision system operates on the principle of triangulation, utilizing the geometric relationship between corresponding points in multiple camera views to estimate depth information [18]. The fundamental mathematical model for stereo vision is based on the pinhole camera model and epipolar geometry principles.

1) Camera model

A 3D object can be projected onto a 2D image using a camera model, most often the pinhole camera. In this model, a point Q in 3D space is converted into a 2D point q on the image plane by the intersection of the line connecting C (camera center) and Q, with the image plane.

In reality, the image is formed behind the camera center. However, for ease of calculation, the positions of the image and the camera center are swapped. The distance of the object from the camera is Z, the object's actual length is X, its image length is x, and the focal length is f. The image length x can be derived using the concept of similar triangles, as shown in equation (1).

$$x = f\frac{X}{Z} \tag{1}$$

The world coordinate system (U, V, W) and camera coordinate system (X, Y, Z) are used to locate object position in 3D space. The forward projection is illustrated in Figure 1. A 3D point in space has the coordinates (X, Y, Z). A 2D point q is the projection of point Q and is located at (x, y). Using Thales' theorem, the equation that relates the world point to the corresponding image point is expressed in equation (2).

$$x = f\frac{X}{Z} \text{ and } y = f\frac{Y}{Z} \tag{2}$$

As these equations are non-linear, homogeneous coordinates are employed for their solution. A 2D point is determined using the coordinates $(\overline{x}, \overline{y}, \overline{z})$ as shown in equation (3).

$$x = \frac{\overline{x}}{\overline{z}} \text{ and } y = \frac{\overline{y}}{\overline{z}} \tag{3}$$

In Figure 1, the projection matrix $M_{projection}$ is depicted, with f representing the focal length. The extrinsic parameters are contained within $M_{extrinsics}$, which include a 3x3 rotation matrix R and a 3x1 translation vector T. These parameters determine the camera's orientation and position relative to the world frame and may vary depending on the specific world.

The matrix M_{affine} is utilized to transform (x, y) to (u, v), pixel coordinates. This transformation considers two factors: the scale factor and the shift of the principal

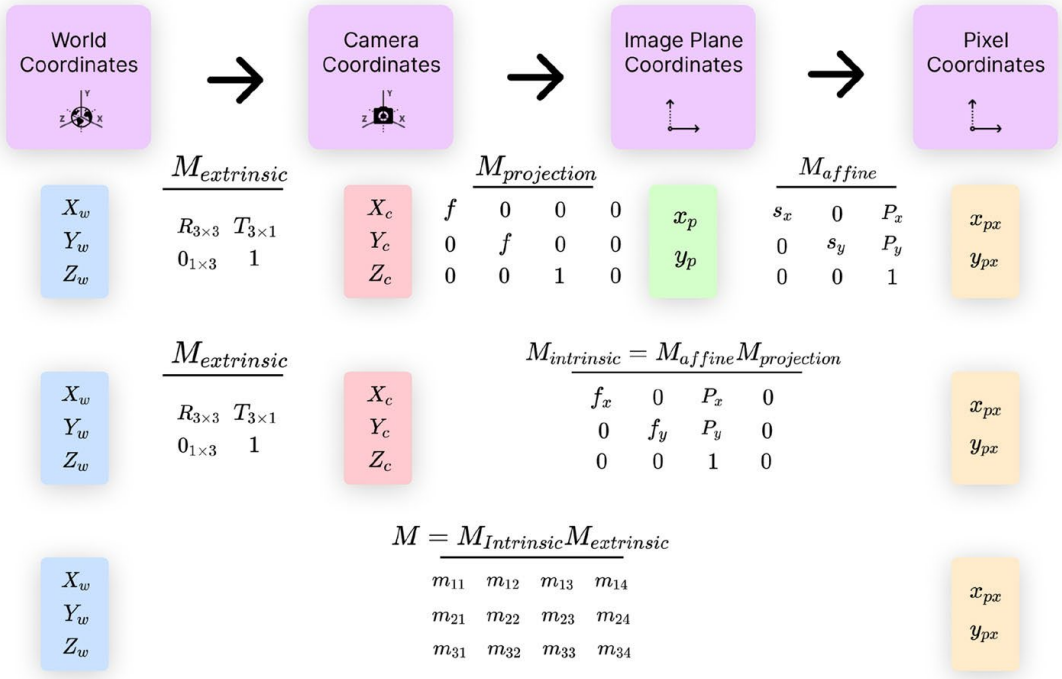

Figure 1. Forward projection [5].

point toward the upper-left corner. The intrinsic parameters are unique to each camera and are contained within the matrix K. These parameters facilitate the conversion between the camera coordinates and the pixel coordinates within the image frame. A displacement of the center of the image plane is denoted by $(p_x, p_y)^T$, while the product of the physical focal length and s_x gives f_x, and f_y is derived from the product of the physical focal length and s_y of the pixel.

2) Lens distortion

In practical application, no lens is perfectly free from distortion. The primary types of lens distortions, which include radial and tangential distortion, are detailed and modeled by Brown [19] and Fryer and Brown [20]. Radial distortion results from the inherent curvature of the lens, whereas tangential distortion arises from imperfections in the camera's assembly or alignment. The initial position of the distorted point is (x, y), and its radial location is determined using equation (4).

$$\begin{pmatrix} x_{cor} \\ y_{cor} \end{pmatrix} = \begin{pmatrix} x(1+k_1r^2+k_2r^4+k_3r^6) \\ y(1+k_1r^2+k_2r^4+k_3r^6) \end{pmatrix} \quad (4)$$

where (x_{cor}, y_{cor}) denotes the new position obtained after distortion. The tangential distortion is represented by two parameters p_1 and p_2 as shown in equation (5).

$$\begin{pmatrix} x_{cor} \\ y_{cor} \end{pmatrix} = \begin{pmatrix} x+(2p_1xy+p_2(r^2+2x^2)) \\ y+(p_1(r^2+2y^2)+2p_2xy) \end{pmatrix} \quad (5)$$

OpenCV was used to fix lens distortion as a part of image processing. OpenCV is an open sources image processing library [21]. OpenCV packages all the five distortion coefficients to form a 1×5 distortion vector in the following order: k_1, k_2, k_3, p_1, and p_2.

3) Stereo geometry and disparity

A basic setup with two cameras to capture an image of the same 3D point Q is shown in Figure 2. The stereo geometry is based on epipolar geometry. The QC_lC_r plane is called the epipolar plane where the point Q is projected into the left and right image planes [22]. These projections are referred to as q_l (left) and q_r (right).

All points along the (C_lQ) form a line, (e_rq_r), on the right image plane. This line is called the epipolar line associated with the q_l point. It is essential to note that the corresponding point of q_l on the right image plane must lie along this line. In simpler terms, one only needs to search along the epipolar line to the corresponding point, rather than searching the entire image. This concept is known as the epipolar constraint. Similarly, every point possesses its corresponding epipolar line in the opposite image.

C_l and C_r represent the left and right camera centers, respectively. As per the configuration depicted in Figure 2, the projection of C_r onto the left image plane is e_l, which is called the left epipole. In an analogous manner, e_r is the right epipole. These epipoles are the points where the image planes intersect with the baseline C_lC_r.

Through the stereo rectification process, acquired images are used to generate the disparity map. As illustrated in Figure 3, the distance separating the

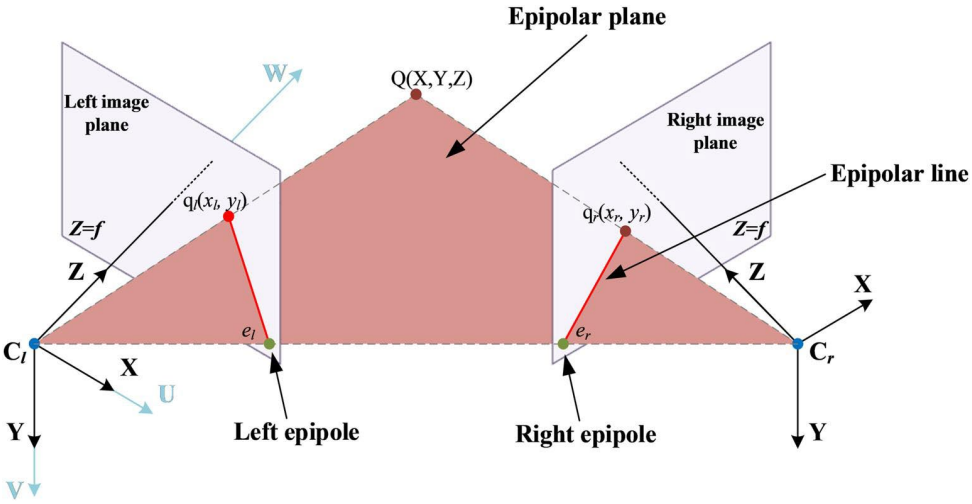

Figure 2. Epipolar geometry: general case [9].

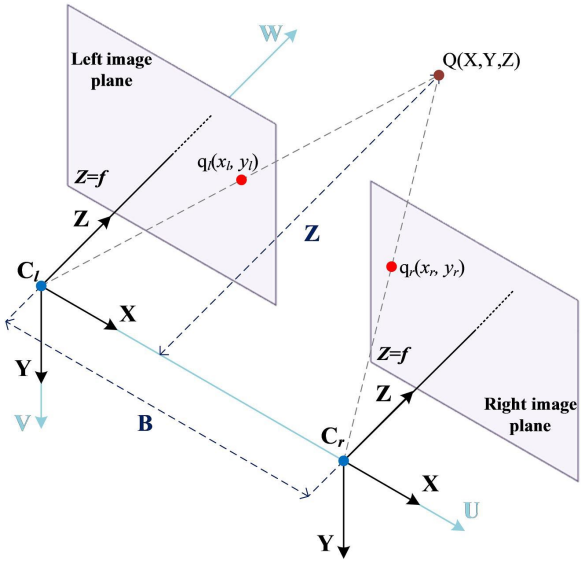

Figure 3. Stereo triangulation scheme [9].

centers of the left and right cameras is denoted as B. The coordinates $q_l(x_l, y_l)$ for the projection of Q onto the left image plane and $q_r(x_r, y_r)$ for the projection onto the right image plane are defined accordingly [22]. The variable Z signifies the distance from Q to the camera. After the stereo rectification, the y coordinates share common values, i.e., $y = y_l = y_r$.

By applying similar triangles, the following equation can be derived as shown in equation (6).

$$\frac{B}{Z} = \frac{B-(x_l-x_r)}{Z-f} \tag{6}$$

The term $x_l - x_r$ is known as disparity. Let $d = x_l - x_r$ and the equation can be rewritten as equation (7).

$$\frac{B}{Z} = \frac{B-d}{Z-f} \rightarrow Z = \frac{f \times B}{d} \tag{7}$$

In equation (7), f and B are fixed values, the only unknown parameter is d. For this reason, it is necessary to determine disparity in order to calculate the distance between the obstacle and the camera.

B. Semi-global block matching algorithm

The recovery of three-dimensional structure from two-dimensional images is a fundamental objective in computer vision. Among the various techniques developed to achieve this, binocular stereo vision stands as one of the most established and powerful. It mimics the human visual system's ability to perceive depth by finding and analyzing the differences between two images of the same scene captured from slightly different viewpoints. The output of this process, a dense disparity map, provides a rich, per-pixel representation of the scene's geometry, enabling a wide range of applications from autonomous navigation to photogrammetric mapping.

By the early 2000s, the field of stereo vision was largely defined by this stark trade-off: the speed of local methods came at the cost of accuracy, while the accuracy of global methods came at the cost of feasibility [23]. This polarization created a significant barrier to the deployment of high-quality, dense stereo vision in practical, resource-constrained systems, such as mobile robots and autonomous vehicles, which demand both high accuracy and real-time performance. The theoretical optimality of global methods proved to be a dead end for applications where low latency was essential.

This gap in the algorithmic landscape created a clear need for a new approach-one that could bridge the divide by combining the strengths of both paradigms. The goal was to develop an algorithm capable of achieving accuracy close to that of global methods but with computational demands similar to those of local methods.

In 2005, Hirschmüller introduced semi-global block matching (SGBM), a novel technique that provided an elegant and practical solution to this long-standing problem. The central idea behind SGBM was to avoid solving the NP-hard 2D global optimization problem directly [24]. Instead, SGBM approximates the minimization of the 2D global energy function by aggregating the results of multiple independent, one-dimensional optimizations.

Each of these 1D optimizations is performed along a path through the image and can be solved efficiently using dynamic programming. By combining the costs from multiple paths passing through each pixel from different directions (e.g., horizontal, vertical, and diagonals), SGBM effectively incorporates a global-like smoothness constraint without incurring the prohibitive computational cost of a true 2D optimization [25]. This clever approximation allowed SGBM to achieve an excellent balance between accuracy and efficiency, delivering high-quality, dense disparity maps with sharp object boundaries at runtimes suitable for practical applications [25]. SGBM's success was not merely an incremental improvement; it represented a new way of thinking about computationally hard problems in vision. It demonstrated that a tractable approximation of a theoretically optimal model could outperform existing practical methods. This established a powerful design pattern and cemented SGBM's status as a foundational "classical" algorithm, leading to its widespread implementation in popular libraries such as OpenCV and on dedicated hardware such as field-programmable gate array (FPGAs) [24].

To fully appreciate the behavior and capabilities of the Semi-Global Matching algorithm, it is essential to first understand the mathematical objective it seeks to achieve. At its core, SGBM is an energy minimization framework. It aims to find a disparity image D that minimizes a global energy function $E(D)$, which is carefully designed to balance the visual evidence from the images with prior knowledge about the geometric structure of the world.

The global energy function $E(D)$ assigns a total cost to a candidate disparity map D. A lower energy corresponds to a more plausible disparity map. The function is defined as a sum over all pixels p in the image and is composed of two fundamental terms: a data term and a smoothness term. In its general form, it can be written as equation (8).

$$E(D) = \sum_p (C(p, D_p) + \sum_{q \in N_p} V(D_p, D_q)) \quad (8)$$

where p and q are pixels, D_p is the disparity assigned to pixel p, $C(p, D_p)$ is the matching cost (the data term) for pixel p at disparity Dp, Np represents the set of neighboring pixels of p, and $V(D_p, D_q)$ is a penalty function (the smoothness term) that penalizes differences in disparity between neighboring pixels p and q.

C. Experimental setup and evaluation metrics

This research follows the design science research methodology framework, which provides a structured approach for developing and evaluating technological artifacts [26]. The experiments were conducted using the CARLA simulator version 0.9.15, configured with specifications as follows: environment - Town10HD; weather - middle of the day; camera setup - setup pair with 20 cm baseline. The camera parameters were configured as follows: baseline distance of 0.2 meters, field of view of 90°, and image resolution of 640 × 480 pixels. Test scenarios were designed to evaluate the algorithm's performance under different vehicle orientations, with the front wheel heading set toward east, north, south, and west relative to the horizon. The evaluated distance range was 40-350 cm.

To assess the algorithm's accuracy, several evaluation metrics were employed, including relative error, absolute error, and root mean square error (RMSE). These metrics were computed for each test scenario within the simulated environment. The finding from this evaluation serves as a validation of the algorithm's effectiveness.

D. Distance estimation model

A distance estimation system based on the SGBM algorithm was developed and implemented on a personal computer running Windows 10, equipped with an Intel Core i9 CPU and 64 GB of RAM, and

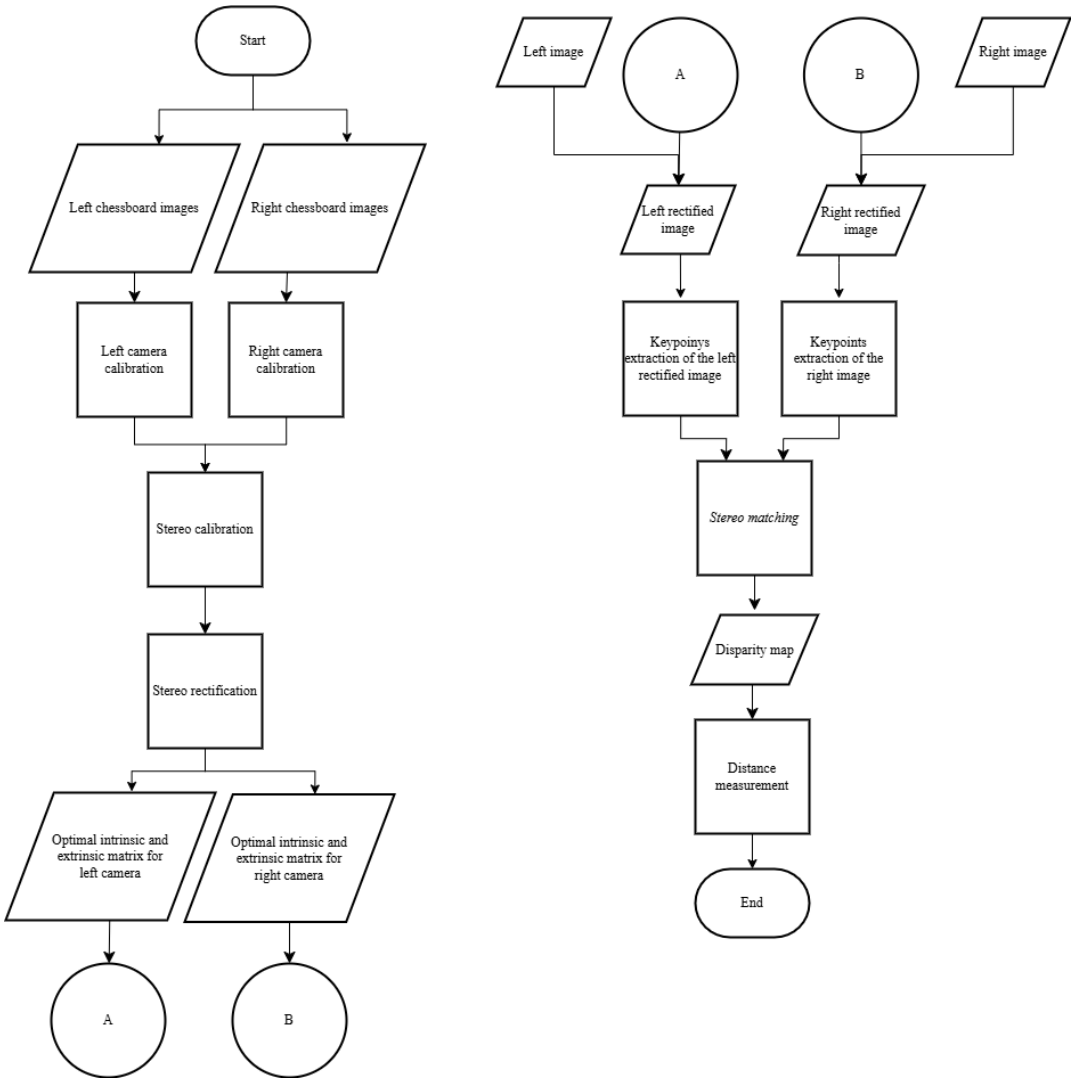

Figure 4. Flowchart of the proposed distance estimation model.

programmed in Python. The stereo vision system consists of two identical RGB cameras with a 90° field of view, both provided within the CARLA simulator environment.

Camera calibration was performed using a 10x7 chessboard pattern. The chessboard was rotated and positioned to ensure that most of the camera's field of view was captured in the collected images. To achieve reliable calibration, at least ten images of the chessboard were required; however, thirty-five images were collected from various orientations to enhance precision. The flowchart of the distance estimation model is illustrated in Figure 4, which can be divided into four primary stages.

The first step of the process is a stereo calibration. In this stage, each camera was calibrated individually by detecting the corners of the chessboard across 35 images. A sub-pixel corner detection method was employed to achieve high precision in corner localization. Subsequently, stereo calibration was performed, producing two 3x4 projection matrices that define the camera in a unified coordinate system.

The second step is stereo rectification. Stereo rectification aligns the left and right images such that corresponding epipolar lines are parallel and horizontally aligned. This process reprojects both camera images onto a common image plane, simplifying the stereo matching task and improving accuracy. The rectified image pair is then used to compute the disparity map in the next stage.

The third step is stereo matching. Stereo matching determines depth information from two or more images of the same scene captured from slightly different viewpoints, similar to how humans' vision processes depth using both eyes. The depth information received is the disparity (horizontal displacement) between points in the left and right images. A commonly used algorithm for stereo matching is semi-global matching [27]. The result of stereo matching is a disparity map that we can use in the distance calculation stage.

The final step is distance calculation using the disparity map obtained from the stereo matching process. The distance is calculated by converting the disparity value at a known distance using a sixth-order polynomial equation.

III. Results and Discussions

Table 1 presents the relationship between disparity values calculated by the stereo vision model and the corresponding measured distances. Using regression analysis, polynomial functions were derived to estimate distance. The polynomial function's values of the 3rd, 5th, and 6th orders are shown in Figure 5.

Based on the polynomial order graph in Figure 5, the sixth-order polynomial regression was selected because it achieved the lowest root mean square error (RMSE) among all tested regression orders, indicating the most accurate approximation of the distance–response relationship. RMSE is widely regarded as a reliable indicator of predictive precision in continuous-value estimation tasks [28][29]. Although higher-order

Table 1.
Relationship between disparity value and the measured actual distance.

Actual distance (cm)	Average disparity (px)	Standard deviation	Actual distance (cm)	Average disparity (px)	Standard Deviation
40	0.03414807	0.00016342	200	-0.00886305	0.00002387
50	0.02522009	0.00008927	210	-0.00960398	0.00006535
60	0.01871172	0.00011413	220	-0.01030389	0.00003941
70	0.01368315	0.00006651	230	-0.01080968	0.00002228
80	0.00972427	0.00013637	240	-0.01129703	0.00002448
90	0.00651247	0.00010491	250	-0.01174746	0.00004608
100	0.00390147	0.00006211	260	-0.01221156	0.00003463
110	0.00162538	0.00009788	270	-0.01271941	0.00007586
120	-0.00019958	0.00006599	280	-0.01303792	0.00010301
130	-0.00186666	0.00003692	290	-0.01333935	0.00004478
140	-0.00335124	0.00003257	300	-0.01374809	0.00002084
150	-0.00447971	0.00004670	310	-0.01398526	0.00014081
160	-0.00556786	0.00005146	320	-0.01422654	0.00003387
170	-0.00646120	0.00004825	330	-0.01450541	0.00007780
180	-0.00738394	0.00000985	340	-0.01473029	0.00001716
190	-0.00832923	0.00005771	350	-0.01504812	0.00013390

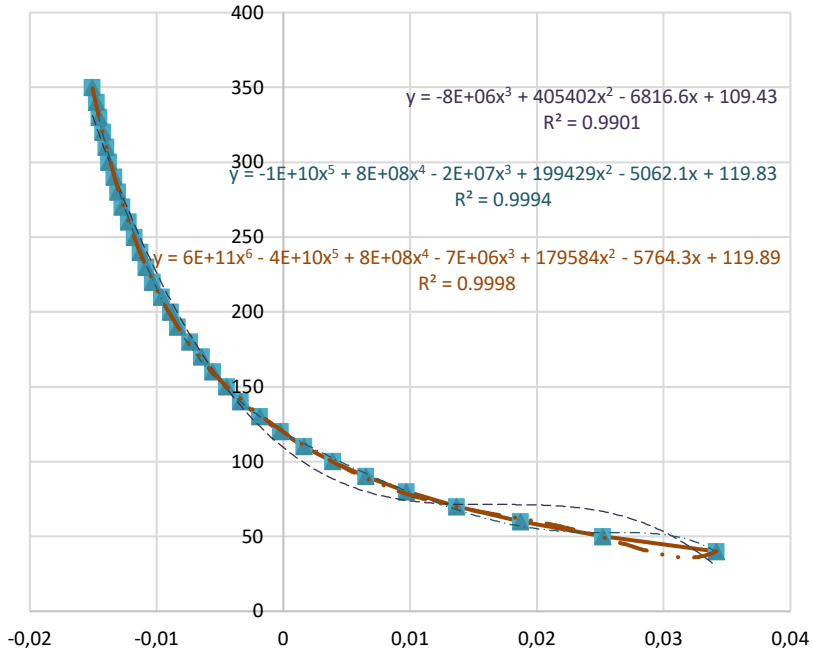

Figure 5. Polynomial graph of 3rd, 5th, and 6th.

Table 2.
The relationship between the estimated and measured distance.

Actual distance (cm)	Estimated distance (cm)	Absolute error (cm)	Relative error (%)	Actual distance (cm)	Estimated distance (cm)	Absolute error (cm)	Relative error (%)
40	35.922	4.078	10.195	200	198.484	1.516	0.758
50	46.362	3.638	7.276	210	210.436	0.436	0.208
60	61.298	1.298	2.163	220	220.634	0.634	0.288
70	69.332	0.668	0.954	230	230.076	0.076	0.033
80	78.296	1.704	2.130	240	239.702	0.298	0.124
90	89.376	0.624	0.693	250	249.674	0.326	0.130
100	99.824	0.176	0.176	260	261.788	1.788	0.688
110	110.432	0.432	0.393	270	270.46	0.46	0.170
120	120.586	0.586	0.488	280	279.266	0.734	0.262
130	131.024	1.024	0.788	290	290.714	0.714	0.246
140	141.620	1.62	1.157	300	304.158	4.158	1.386
150	150.268	0.268	0.179	310	308.974	1.026	0.331
160	160.142	0.142	0.089	320	319.574	0.426	0.133
170	168.924	1.076	0.633	330	330.018	0.018	0.005
180	178.292	1.708	0.949	340	338.48	1.52	0.447
190	188.608	1.392	0.733	350	354.25	4.25	1.214

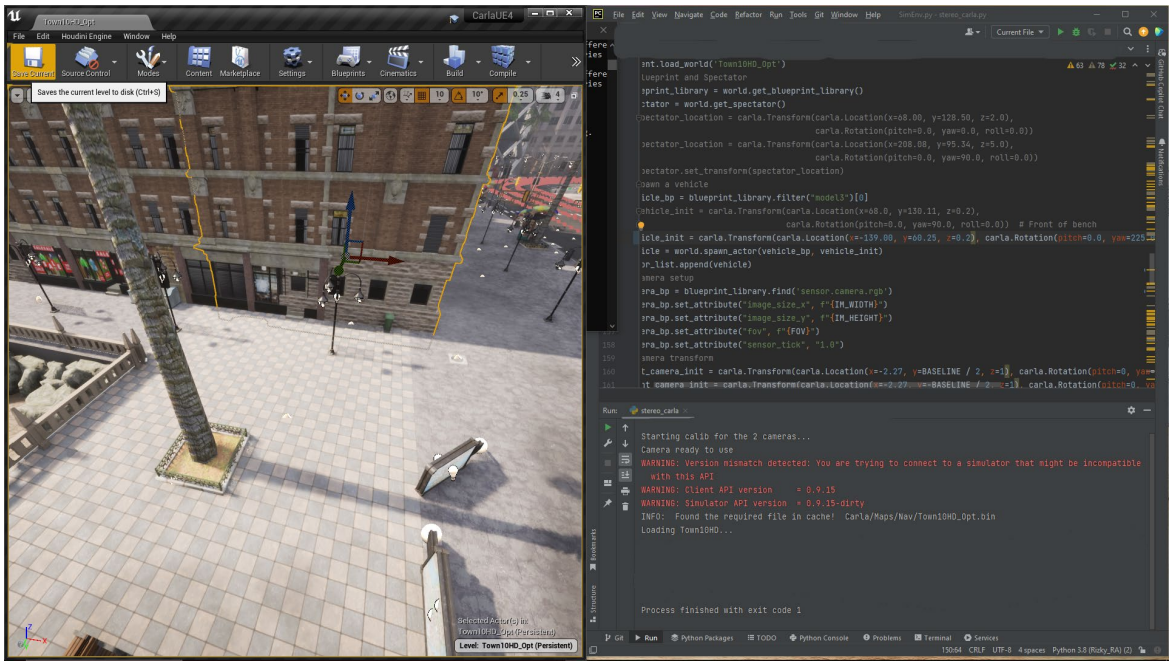

Figure 6. Ilustration of promenade and the simulator environment configuration.

polynomials may introduce unnecessary complexity or oscillatory behavior, an examination of the residual distribution and cross-validated performance confirmed that the sixth-order model provided improved accuracy without evidence of overfitting. This decision is consistent with recent findings in the literature, where modern studies emphasize minimizing prediction error while applying appropriate regularization or model-selection criteria. Zhao *et al.* [30] demonstrated the importance of basis-function selection and cross-validation for achieving optimal polynomial response surfaces, while Zhang *et al.* [31] showed that information-driven fitting methods—evaluated using RMSE and information criteria—yield improved accuracy and stability in sensor-related regression tasks.

Table 2 presents the experiment results comparing the estimated distances produced by the model with the actual distances. The actual distance is based on the coordinates in CARLA. Figure 6 presents that every

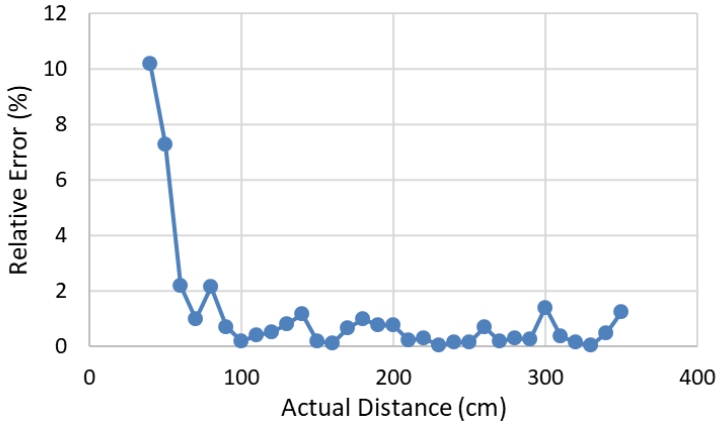

Figure 7. Relative error.

1 unit of coordinate (x, y and z) is 1 meter in distance. The coordinate was declared by code, as a part of CARLA's simulator environment configuration. Based on the data, the RMSE was calculated to be 1.69 cm and the average relative error is below 1.1 % in the distance range of 60 cm - 350 cm. This performance is comparable to results reported in state-of-the-art studies in this field, which generally record relative errors ranging from 0.1 % to 2.2 % [22]. The strong performance of the proposed model is also attributed to the precise stereo calibration process, which achieved an RMSE below 0.5 pixels, a commonly accepted threshold for high-quality calibration.

For distances below 80 cm, the distance estimation model demonstrates high inaccuracy, as illustrated in Figure 7. The experimental result shows a large relative error exceeding 10 % at the closest point, indicating that the model is unreliable for short-range distance estimation. Once the actual distance surpasses approximately 80-100 cm, the model's performance improves significantly. The relative error stabilizes and remains consistently low, generally fluctuating within a narrow range between 0 % and 2 %. This range represents the model's optimal operating region, where it provides the most reliable and accurate estimations.

IV. Conclusion

This research evaluated the SGBM algorithm for distance estimation in an autonomous parking application using the CARLA simulator. The experiment results demonstrate the effectiveness of the proposed distance estimation model, achieving an overall RMSE of 1.69 cm and an average relative error of 1.1 % across the critical distance range of 40-350 cm. Future research directions include extending the evaluation to additional environmental conditions, investigating the integration of machine learning techniques for further accuracy improvements, and conducting real-world validation. Additionally, exploration of multimodal sensor fusion combining stereo vision with other sensing modalities could provide enhanced robustness and accuracy for autonomous parking applications.

Acknowledgements

The authors would like to thank the National Research and Innovation Agency (BRIN) for supporting the project with resources. The authors wish to express their sincere gratitude to Mr. Aryo Putro Sadono and Mr. Asep Nugroho for their invaluable support in preparing the computational workstation used for the simulations. We are also grateful to Mr. Oka Mahendra for the guidance given to the first author when applying to the Research Assistantship program at BRIN.

Declarations

Author contribution

R. H. Sakti and E. Rijanto are the main contributors of this article. All authors read and approved the final paper.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-forprofit sectors.

Conflict of interest

The authors declare no conflict of interest.

The use of AI or AI-assisted technologies

During the preparation of this work, the author(s) used Perplexity in order to find the novelty of research. After using this tool/service, the author(s) reviewed

and edited the content as needed and take(s) full responsibility for the content of the publication.

References

[1] J. Toledo, M. Lauer, and C. Stiller, "Real-time stereo semi-global matching for video processing using previous incremental information," *J. Real-Time Image Process.*, vol. 19, no. 1, pp. 205–216, Feb. 2022.

[2] S. Shabani, A. R. Mulaka, T. A. Stokes, T. U. Rehman, and Y. Bao, "Development of a stereo vision-based UGV guidance system for bareroot forest nurseries," *Smart Agric. Technol.*, vol. 11, p. 100990, Aug. 2025.

[3] Y. Chen, F.-Y. Huang, B.-Q. Liu, S. Zhang, Z. Wang, and B. Zhao, "Significant obstacle location with ultra-wide FOV LWIR stereo vision system," *Opt. Lasers Eng.*, vol. 129, p. 106076, June 2020.

[4] D. Fernandes et al., "Real-time 3D object detection and SLAM fusion in a low-cost LiDAR test vehicle setup," *Sensors*, vol. 21, no. 24, p. 8381, Dec. 2021.

[5] Y. Wang, Z. Jiang, Y. Li, J.-N. Hwang, G. Xing, and H. Liu, "RODNet: A real-time radar object detection network cross-supervised by camera-radar fused object 3D localization," *IEEE J. Sel. Top. Signal Process.*, vol. 15, no. 4, pp. 954–967, June 2021.

[6] Y. Jo, J. Ha, and S. Hwang, "Survey of technology in autonomous valet parking system," *Int. J. Automot. Technol.*, vol. 24, no. 6, pp. 1577–1587, Dec. 2023.

[7] M. Elsayed, M. T. Soe, W. W. Kit, and H. Abdalla, "An innovative approach to developing a 3D virtual map creator using an ultrasonic sensor array," *Int. J. Technol.*, vol. 10, no. 7, p. 1344, Nov. 2019.

[8] C.-C. Lo and P. Vandewalle, "Expanding sparse radar depth based on joint bilateral filter for radar-guided monocular depth estimation," *Sensors*, vol. 24, no. 6, p. 1864, Mar. 2024.

[9] W. Hua et al., "Key technologies in apple harvesting robot for standardized orchards: A comprehensive review of innovations, challenges, and future directions," *Comput. Electron. Agric.*, vol. 235, p. 110343, Aug. 2025.

[10] M. Grabowski and T. Kryjak, "Real-time FPGA implementation of the semi-global matching stereo vision algorithm for a 4K/UHD video stream," in *Design and Architecture for Signal and Image Processing*, M. Chavarrías and A. Rodríguez, Eds., Cham: Springer Nature Switzerland, 2023, pp. 70–81.

[11] S. Hermann and R. Klette, "Iterative semi-global matching for robust driver assistance systems," in *Computer Vision – ACCV 2012*, vol. 7726, K. M. Lee, Y. Matsushita, J. M. Rehg, and Z. Hu, Eds., in Lecture Notes in Computer Science, vol. 7726, Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 465–478.

[12] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "CARLA: An open urban driving simulator," in *Proceedings of the 1st Annual Conference on Robot Learning*, 2017, pp. 1–16.

[13] Y. Ye, Y. Long, W. Li, and Y. Yi, "Depth-driven machine vision framework for accurate vehicle speed measurement in dynamic environments," *Measurement*, vol. 253, p. 117717, Sept. 2025.

[14] T. Guan, J. Cai, Y. Wang, W. Yang, X. Chang, and Y. Han, "Pavement pothole detection system based on deep learning and binocular vision," *J. Traffic Transp. Eng. Engl. Ed.*, vol. 12, no. 4, pp. 1100–1123, Aug. 2025.

[15] Z. Zhu, J. Zhu, and J. Wei, "Autonomous vehicle system based on A* pathfinding and improved YOLOv11," in *Proceedings of the 2025 2nd International Conference on Autonomous Driving and Intelligent Sensing Technology*, Guilin China: ACM, Mar. 2025, pp. 34–39.

[16] G. Heng, C. Luo, K. Sun, S. Huang, and L. Huang, "SGBM_YOLO: A high-precision obstacle detection algorithm for assistive navigation based on stereo vision and spatial attention mechanism," *Appl. Soft Comput.*, vol. 186, p. 114144, Jan. 2026.

[17] A. D. Taromi and S. H. Klidbary, "A novel data-driven algorithm for object detection, tracking, distance estimation, and size measurement in stereo vision systems," *Multimed. Tools Appl.*, vol. 84, no. 12, pp. 11041–11061, May 2024.

[18] M. N. Flinders, P. Rao, A. R. Reibman, D. R. Buckmaster, and J. P. Boerman, "Design and validation of a stereo vision system to estimate volume of total mixed rations offered to dairy cattle," *Comput. Electron. Agric.*, vol. 238, p. 110816, Nov. 2025.

[19] D. Brown. (1971). *Close-range camera calibration Photo.* Engng. [Online].

[20] J. G. Fryer and D. C. Brown, "Lens distortion for close-range photogrammetry," *Photogramm. Eng. Remote Sens.*, vol. 52, pp. 51–58, Jan. 1986.

[21] S. Sivkov et al., "The algorithm development for operation of a computer vision system via the OpenCV library," *Procedia Computer Science*, vol. 169, pp. 662–667, Jan. 2020.

[22] E. Adil, M. Mikou, and A. Mouhsen, "A novel algorithm for distance measurement using stereo camera," *CAAI Trans. Intell. Technol.*, vol. 7, no. 2, pp. 177–186, June 2022.

[23] Y.-Q. Guo, M. Gu, and Z.-D. Xu, "Research on the improvement of semi-global matching algorithm for binocular vision based on lunar surface environment," *Sensors*, vol. 23, no. 15, p. 6901, Aug. 2023.

[24] H. Hirschmüller, "Semi-global matching - motivation, developments and applications," in *Photogrammetric Week 11*, D. Fritsch, Ed., Wichmann, Sept. 2011, pp. 173–184. [Online].

[25] E. Dall'Asta and R. Roncella, "A comparison of semiglobal and local dense matching algorithms for surface reconstruction," *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. XL–5, pp. 187–194, June 2014.

[26] D. T. Absari, A. Djunaidy, and T. D. Susanto, "Design science research methodology concept and its application in the development of online community communication system for Selotapak village," in *2022 9th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)*, Semarang, Indonesia: IEEE, Aug. 2022, pp. 236–241.

[27] H. Hirschmuller, "Accurate and efficient stereo processing by semi-global matching and mutual information," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, June 2005, pp. 807–814 vol. 2.

[28] A. Panariello, G. Mancusi, F. Haj Ali, A. Porrello, S. Calderara, and R. Cucchiara, "Monocular per-object distance estimation with masked object modeling," *Comput. Vis. Image Underst.*, vol. 253, p. 104303, Mar. 2025.

[29] S. G. Pundkar and A. Gaikwad, "Search and hunt based-driving scene perception network for distance estimation and object detection," *Digit. Signal Process.*, vol. 149, p. 104506, June 2024.

[30] Y. Zhao, S. Ye, X. Chen, Y. Xia, and X. Zheng, "Polynomial response surface based on basis function selection by multitask optimization and ensemble modeling," *Complex Intell. Syst.*, vol. 8, no. 2, pp. 1015–1034, Apr. 2022.

[31] W.-H. Zhang, L. Dai, W. Chen, A. Sun, W.-L. Zhu, and B.-F. Ju, "Novel information-driven smoothing spline linearization method for high-precision displacement sensors based on information criterions," *Sensors*, vol. 23, no. 22, p. 9268, Nov. 2023.