



Article

A Vehicle Monocular Ranging Method Based on Camera Attitude Estimation and Distance Estimation Networks

Jun Liu * and Duo Xu

School of Automotive and Transportation Engineering, Jiangsu University, Zhenjiang 212013, China; 2222104029@stmail.ujs.edu.cn

* Correspondence: liujun@ujs.edu.cn

Abstract: A monocular ranging method for forward vehicles in intelligent driving is proposed. This method measures vehicle distance more accurately under the condition of a single camera and can estimate camera attitude in real-time. For the estimation of camera pitch and yaw angles, it is achieved using road vanishing points. The images collected by the camera are sequentially processed through the Roberts operator amplitude calculation, feature point extraction, feature line segment generation, road vanishing point voting, and estimation of camera attitude to obtain pitch and yaw angles. A distance estimation network was designed, which is divided into multiple levels based on image size and incorporates image feature, integrating vehicle grounding point, and vehicle width information, effectively improving ranging accuracy. Finally, validation was conducted on KITTI data, with a relative error (AbsRel) of 8.3%. Additionally, the TuSimple dataset and continuous driving scenarios were also validated, resulting in improved performance compared to previous algorithms.

Keywords: machine vision; monocular; vehicle ranging; camera pose; ranging network



Citation: Liu, J.; Xu, D. A Vehicle Monocular Ranging Method Based on Camera Attitude Estimation and Distance Estimation Networks. *World Electr. Veh. J.* **2024**, *15*, 339. <https://doi.org/10.3390/wevj15080339>

Academic Editor: Michael Fowler

Received: 18 June 2024

Revised: 15 July 2024

Accepted: 24 July 2024

Published: 27 July 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In the process of implementing autonomous driving, Computer Vision (CV) plays an important role, including vehicle ranging, vehicle speed measurement, vehicle detection, and traffic sign recognition [1]. Among these, distance perception is a key technology for traffic safety and path planning. With the continuous improvement in visual intelligence, vehicle ranging has significantly improved in accuracy and efficiency [2].

Demonstrating the importance of this solution, monocular distance measurement for vehicles ahead can be applied to assist driving safety, autonomous driving, and has a simple structure that is easy to implement. Compared to radar, lidar, stereo-cameras, etc., monocular cameras are more economical and suitable for vehicles with limited budgets. In addition, the monocular camera has a compact structure and is easier to install in different positions of the vehicle. The data processing of monocular cameras is usually relatively simple and requires low computing resources.

Due to the complexity of vehicle driving under various road conditions, the images captured by cameras are influenced by various factors, such as posture changes, rain and snow, etc. Therefore, the effectiveness of camera image acquisition significantly affects the subsequent vehicle detection and distance measurement. At present, the main theories for camera pose correction include those based on six lane points [3], road disappearance points [4], virtual horizontal lines [5], feature point motion [6], etc. Among these, correcting the camera pose based on road disappearance points can effectively improve measurement accuracy.

The traditional distance measurement methods for vehicles ahead include those based on the contact point between the vehicle and road [7], and the estimated information of the vehicle target [5]. The method of using ground point distance measurement is greatly affected by road bumps due to its reliance on camera angle parameters [8]. Because the width of the vehicle ahead is unknown, ranging based on vehicle width often requires

estimating the vehicle width based on the type of vehicle [9], and the accuracy of vehicle width prediction has a significant impact on the ranging effect.

This article proposes a monocular distance measurement method for forward vehicles, which includes camera attitude estimation and a multi-reference distance estimation network. It was validated on the KITTI 2012 dataset [10] and showed improvement compared to previous algorithms.

For the camera pose estimation, the specific process includes edge feature point extraction, edge feature segment generation, road disappearance point voting, and camera angle estimation. The estimation of pose improves the accuracy of subsequent distance measurement.

For the network ranging method with grounding point and vehicle width, a network with multiple image size levels and contour information was designed to address the shortcomings of traditional ranging methods. Combined with geometric ranging methods, the accuracy of distance estimation was effectively improved.

2. Camera Pose Estimation

2.1. Road Disappearance Point Detection

During the driving of vehicles, due to the dynamic changes in road conditions and the bumps and undulations of the vehicle itself, the heading angle and pitch angle of the camera will change, and this uncertain parameter causes errors in vehicle distance measurement. This article uses road disappearance points to correct the heading and pitch angles of the camera and improve ranging accuracy.

Firstly, in order to obtain the disappearance point of the road, it is necessary to extract the features of the image. In order to obtain image features with reasonable time complexity, a texture (gradient)-based algorithm is chosen to obtain edge feature information of the image.

Common edge extraction operators include Roberts, Prewitt, Sobel, and Laplacian. KITTI (2012) is used as the dataset, which is converted into grayscale images, and the accuracy of lane line extraction is obtained with a threshold of 50%. The accuracy of lane line extraction will be greatly improved during the subsequent line segment filtering process. Currently, the accuracy of each operator's extraction is shown in the Figure 1.

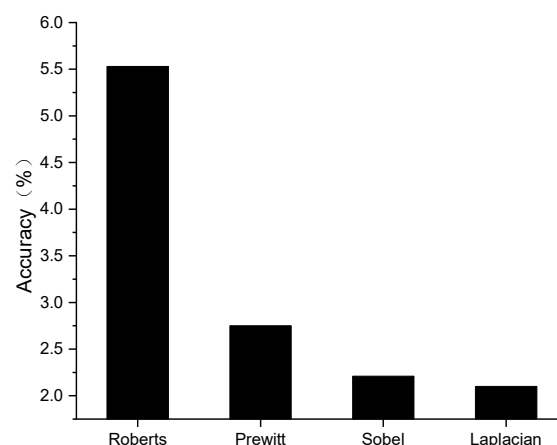


Figure 1. The accuracy of each edge extraction operator in extracting lane lines.

As shown in Figure 1, compared to other operators, the Roberts operator has a significantly higher efficiency in extracting lane lines. In traffic scenarios, lane markings are an obvious feature, and complex operators extract a large number of image detail elements, which results in low efficiency in extracting simple features. The Roberts operator itself is a relatively simple operator that can effectively extract the obvious feature of lane lines. The calculation formula for the Roberts operator with a size of 2×2 is as follows:

$$\begin{cases} G(x,y) = \sqrt{G_x^2(x,y) + G_y^2(x,y)} \\ G_x(x,y) = \frac{1}{2} \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix} \times I(x,y) \\ G_y(x,y) = \frac{1}{2} \begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix} \times I(x,y) \end{cases} \quad (1)$$

where $G(x,y)$ is the gradient amplitude of the image, and $G_x(x,y)$ and $G_y(x,y)$ are the components of the gradient on the X and Y axes.

For lane markings, the segment information is more important, and the extraction results of the segment are shown in Figure 2c. In order to improve the efficiency of line segment generation, the least squares method is used for fitting. Firstly, traverse all extracted feature points and preliminarily fit a large number of small line segments. Comparing all small segments, if the difference in slope and intercept between the two is within 5%, it is considered that they are unified segment information and replaced with extended segments. Repeatedly repeating the pairwise merging process of line segments ultimately obtains the preliminary line segment features of the image.

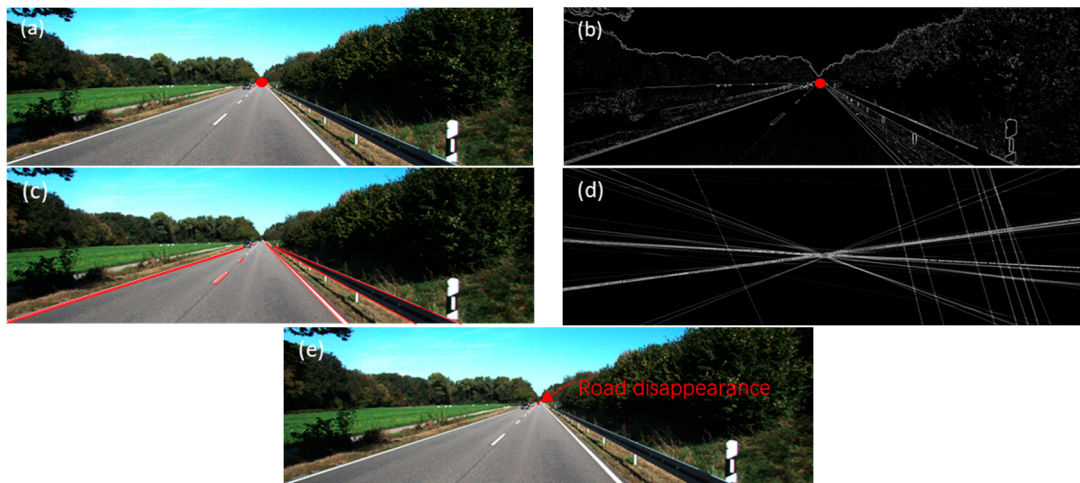


Figure 2. In road vanishing point detection, the results of each step: (a) the original image, No. 11 of the KITTI test set; (b) edge feature extraction; (c) filtered feature line segments; (d) road vanishing point voting; (e) road vanishing point detection results. (The red line is the characteristic line of the road, and the red dot is the point where the road disappears.)

Although the feature line segments of the image have been obtained, there are also a large number of useless line segments in addition to the lane lines. In traffic scenarios, lane lines have distinct characteristics, and useless line segment features can be filtered according to the following principles:

1. In traffic scenarios, the slope of lane markings has certain characteristics. During driving, features such as road signs, greenery, and buildings that are close to vertical can be filtered. In addition, due to the fact that the characteristic lines of shadows and other shadows generated by lighting are mostly horizontal, they can be filtered. In practical operation, line segments smaller than 5° and larger than 85° are filtered to improve the accuracy of feature line segments.
2. During normal driving, the road is not undulating and the lane markings are located in the lower half of the image. Therefore, the line segments in the upper half of the image can be filtered.
3. The lane markings are clearly bright colors, with green indicating road greening in the bright colors. Therefore, the line segments with dark and green background colors can be filtered to improve the accuracy of feature line segments.

As shown in Figure 2d, based on the extracted features of the line segments, the position of the disappearance point on the road is determined by voting. The weight parameters for voting adopt the following length and direction criteria:

- Segment length

In driving scenes, lane lines, as obvious features, often run through the entire picture, meaning that lane lines are often longer feature segments. So, the longer the line segment, the higher the score in the voting process at the disappearance point of the road.

- Segment direction

During the process of driving along the lane line, in the image, the angle of the lane line is close to $45^\circ/135^\circ$ (slope ± 1). After normalization, the direction of the line segment can be obtained.

The calculation formula for voting on road disappearance points can be obtained by combining the length weight and direction of the line segments as follows:

$$\begin{cases} W = W_L + W_O \\ W_L = \frac{\lambda_L L_x}{L_{max}} \\ W_O = \lambda_O \frac{|k_x|}{2k_x^2 + 1} \end{cases}, \quad (2)$$

where W is the vote obtained at the disappearance point of the road; W_L and W_O represent the length and direction of the vote; L_x is the length of a certain feature line segment; L_{max} is the length of the longest feature line segment; λ_L is the voting weight of the length of the line segment; k_x is the slope of the line segment; and λ_O is the voting weight for the length of the line segment.

Furthermore, for the voting results, there are often multiple peaks near the vanishing point, all of which have some degree of deviation. By using Gaussian distribution, the voting space can be smoothed.

Finally, the disappearance point of the road was obtained, as shown in Figure 2e.

As shown in Figure 3, the road vanishing point detection proposed in this article mainly relies on the main feature of the images. Therefore, it has good performance when driving straight during the day and at night. For curved driving, the algorithm in this article can only predict the extension direction of nearby roads. Although it is difficult to accurately predict the disappearance point of a road on a bend, it can still rely on the trend of nearby lanes to provide guidance for camera attitude correction.

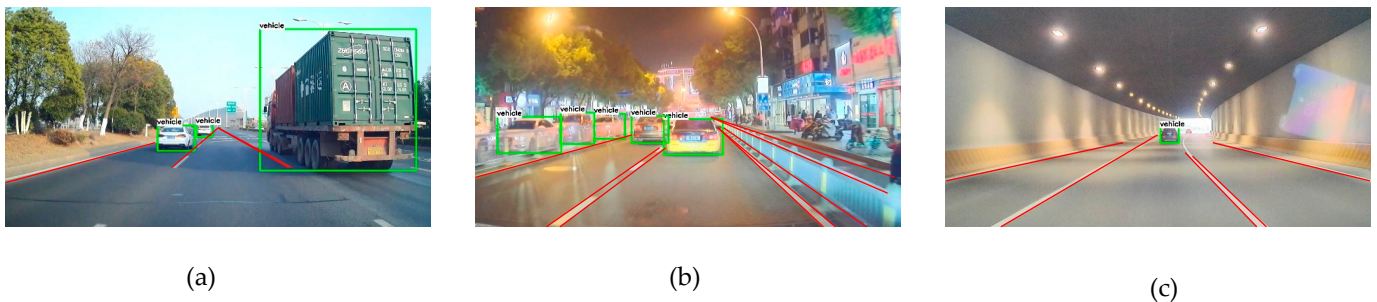


Figure 3. Road vanishing point detection: (a) day; (b) night; (c) curve (The red line is the characteristic line of the road).

2.2. Camera Pose

After determining the location of the disappearance point on the road, the camera's yaw and pitch angles can be estimated based on the offset of the disappearance point's position. After the previous calculation, the coordinates of the disappearance point P on the captured image are $P(u_p, v_p)$, and the coordinates of the camera's optical axis and the initial position O on the captured image are $O(u_o, v_o)$.

As shown in Figure 4b, the yaw angle is γ . After determining the disappearance point of the road, the yaw angle can be calculated based on the initial position of the disappearance point and the camera's optical axis, combined with the camera's parameter f_x . Similarly, as shown in Figure 4a, the camera pitch angle is the angle in the image θ . After determining the location of the disappearance point on the road, it can be obtained based on the relationship shown in the diagram. Based on the disappearance point of the road and the initial position of the camera's optical axis, combined with the camera's parameter f_y , the calculation formula for the camera's yaw and pitch angles is as follows:

$$\begin{cases} \gamma = \tan^{-1}\left(\frac{u_p - u_O}{f_x}\right) \\ \theta = \tan^{-1}\left(\frac{v_p - v_O}{f_y}\right) \end{cases} \quad (3)$$

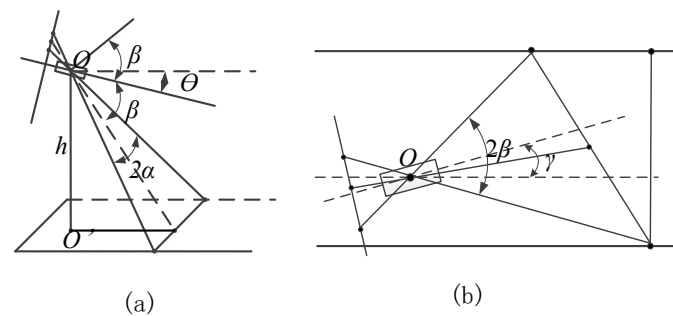


Figure 4. Schematic diagram of yaw angle and pitch angle: (a) pitch angle; (b) yaw angle.

3. Monocular Ranging of the Forward Vehicle

3.1. Distance Measurement Based on Vehicle Grounding Point and Vehicle Width

Due to the lack of degrees of freedom in monocular cameras, external references must be introduced for distance measurement, and the longitudinal position of the front vehicle in the image is a feasible solution. When the camera height is known, the distance between the front vehicle and the self vehicle is shown in Figure 5a.

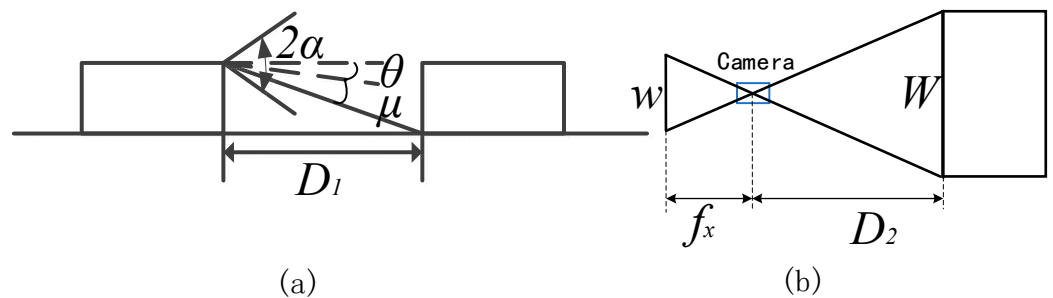


Figure 5. The monocular distance measurement (a) based on the contact point; (b) based on the estimated width.

As shown in Figure 5a, the formula for calculating the distance D_1 between the preceding vehicle and the self-driving vehicle based on the grounding point of the preceding vehicle is as follows:

$$\begin{cases} D_1 = \frac{h}{\tan(\mu)} \\ \mu = \tan^{-1}\left(\frac{v_C - v_O}{f_y}\right) \end{cases} \quad (4)$$

where h is the height of the car-mounted camera; optical center $C(u_C, v_C)$, μ is the angle between the vehicle optical center line and the vertical direction of the optical axis and is calculated by calculating the pixel difference value on the collected image.

The distance measurement scheme based on vehicle width is less affected by the camera attitude angle, but depends on the accuracy of the estimated vehicle width. For the front vehicle, intelligent recognition algorithms can roughly determine the type of front vehicle and estimate the width of the vehicle based on its type. When the width of the vehicle is known, the distance D_2 between the preceding vehicle and the self-driving vehicle can be measured. The specific scheme is shown in Figure 5b, and the distance measurement formula based on the width estimation of the preceding vehicle is as follows:

$$D_2 = \frac{f_x W}{w} \tag{5}$$

where W represents the estimated width of the front car, and w represents the pixel width of the front car.

3.2. The Influence of Attitude Angle on Distance Measurement

The pitch angle will have a certain impact on the ranging results, and the specific impact process is shown in Figure 6a.

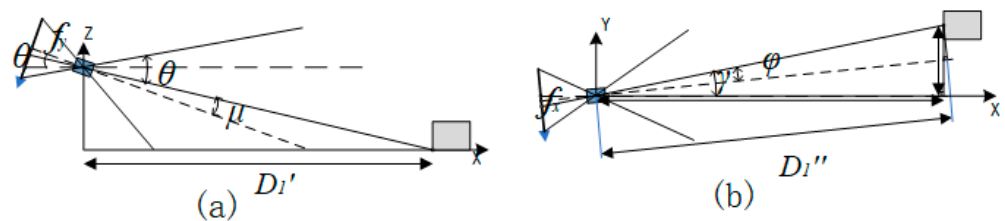


Figure 6. The influence of pitch and yaw angles on ranging results: (a) pitch angle; (b) yaw angle (The blue block represents the camera, and the gray block represents the vehicle being measured).

According to Figure 6, the distance measurement D_1 based on the grounding point should be adjusted. θ represents the camera elevation angle, and D_1' is the distance measurement result adjusted based on the elevation angle of the grounding point.

$$D_1' = \frac{\tan \mu}{\tan(\theta + \mu)} D_1 \tag{6}$$

Similarly, the yaw angle also has a certain impact on the ranging results, as shown in Figure 6b. Due to the yaw angle, D_1' is only the distance from the horizontal projection point of the vehicle on the optical axis. The true distance D_1'' can be calculated according to the following equation:

$$\begin{cases} D_1'' = \frac{D_1'}{\cos \varphi \cdot \cos(\varphi + \gamma)} \\ \varphi = \tan^{-1} \left(\frac{u_c - u_0}{f_x} \right) \end{cases} \tag{7}$$

where φ is the angle between the vehicle optical center line and the horizontal direction of the optical axis.

3.3. The Fusion of Two Ranging Schemes

Two ranging results were obtained based on the grounding point and vehicle width. In order to improve ranging accuracy, it is necessary to design appropriate parameters, integrate two ranging schemes, and obtain the final ranging result D as follows:

$$D = \omega D_1'' + (1 - \omega) D_2, \tag{8}$$

ω represents the weights of two ranging schemes. In order to determine their specific values, the deep learning method was chosen. The network structure and parameters are shown in Figure 7 and Table 1.

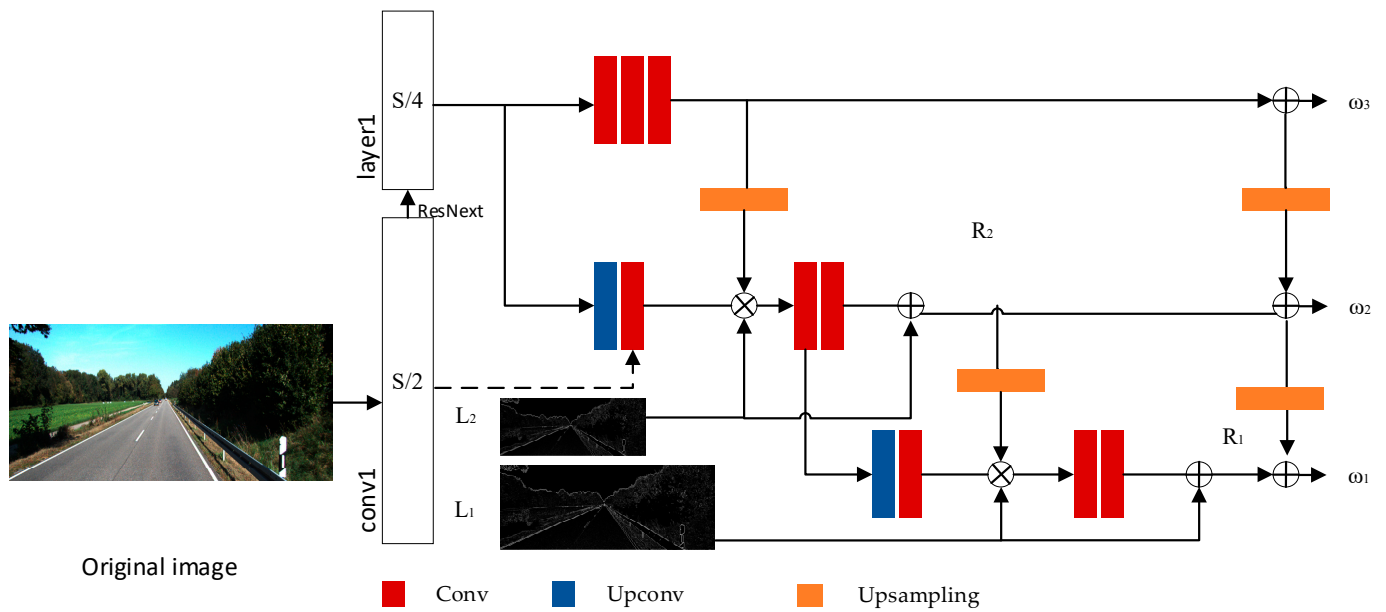


Figure 7. Weighted parameter training network.

Table 1. Detailed structure of the network.

Block	Filter Size	Channel	Input
Encoder			
conv1	7 × 7	3/64	Original image
maxpool	3 × 3	64/64	F (conv1)
layer1	3 × 3	64/256	F (maxpool)
Decoder			
reduction	1 × 1	256/128	F (layer1)
dec3	3 × 3	128/1	F (reduction)
dec2up	3 × 3	128/64	F (dec3bneck)
dec2reduc	1 × 1	128/60	F (conv1 ⊕ dec2up)
dec2bneck	3 × 3	64/64	F (dec2reduc ⊕ R2 ⊕ dec3 *)
dec2	3 × 3	64/1	F (dec2bneck)
dec1up	3 × 3	64/60	F (dec2bneck)
dec1bneck	3 × 3	64/64	F (dec1up ⊕ R1 ⊕ dec2 *)
dec1	3 × 3	64/1	F (dec1bneck)

* for upsampling results, ⊕ for summation.

As shown in Figure 8, during the weight determination process, a parameter network is trained using dataset images. For a certain ranging image, a unique corresponding weight parameter is calculated based on the network to improve ranging accuracy.

The network contains multiple sizes of image inputs, with S/2 and S/4 images being upsampled once and twice, respectively. At the same time, when designing the network, the contour information of the images was taken into consideration, with L1 and L2 as inputs for image contours of different sizes. For the design of the loss function, MSE Loss was adopted.

Images of different sizes focus on features of different sizes; adding more sizes can help improve ranging accuracy, but excessive size input increases network complexity and reduces computational efficiency. The network is divided into multiple levels, with multiple images of different sizes input. When the input is S/4 or S/2, the distance measurement for the preceding vehicle has good accuracy. Different scales of images are inputted to calculate different ranging results, and as the accuracy of the images improves, the ranging

of vehicles ahead becomes more accurate. In Figure 7, ω_1 , ω_2 , and ω_3 correspond to different size inputs, and their ranging accuracy is shown in Figure 8 (using KITTI test set 11 as an example):

In the process of distance measurement, the network introduces the features L_1 and L_2 of the image for training. By introducing the image contour, the training efficiency of the weight network has been significantly improved. It can be seen that the contour of the image during driving is very important information, which is crucial for the design of the training network.

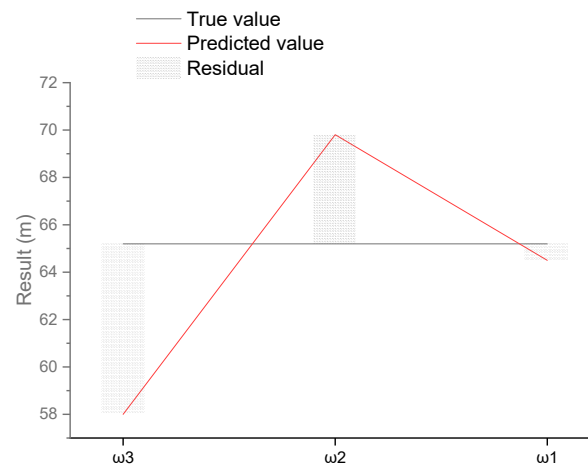


Figure 8. Distance measurement results for multi-scale image input.

4. Results

4.1. Ablation Experiment

The ranging model in this article was validated on two datasets (KITTI and TuSimple), and the prediction results in continuous scenes were tested based on real vehicle, proving that the model has good ranging ability.

As shown in Figure 9, in the distance measurement process of the vehicle ahead, attitude angle estimation and a distance measurement network were introduced. In order to verify the effectiveness of these two modules, ablation experiments were designed. The ablation experiment was based on the KITTI dataset and compared the results of our model, Camera pose estimation, Ranging based on ground point, and Ranging based on vehicle width ranging. The specific values are shown in Table 2.

Table 2. Verification of modules.

Module	(a)	(b)	(c)	(d)
Ranging based on grounding point	✓	✓		✓
Ranging based on vehicle width	✓		✓	✓
Camera pose estimation	✓	✓	✓	
Training network	✓			✓
MSE ↓	1.23	—	—	1.51
Accuracy ↑	0.917	0.853	0.811	0.892

↑ for higher value and better performance, ↓ for lower value and better performance, — for meaninglessness, and ✓ for usage of blocks.

From the table, it can be observed that compared to traditional ground point or vehicle width ranging, our model introduces camera pose estimation and weight training networks, resulting in a significant improvement in ranging accuracy. The accuracy (relative accuracy, complementary to AbsRel) reaches 91.7%, and the loss is relatively low.

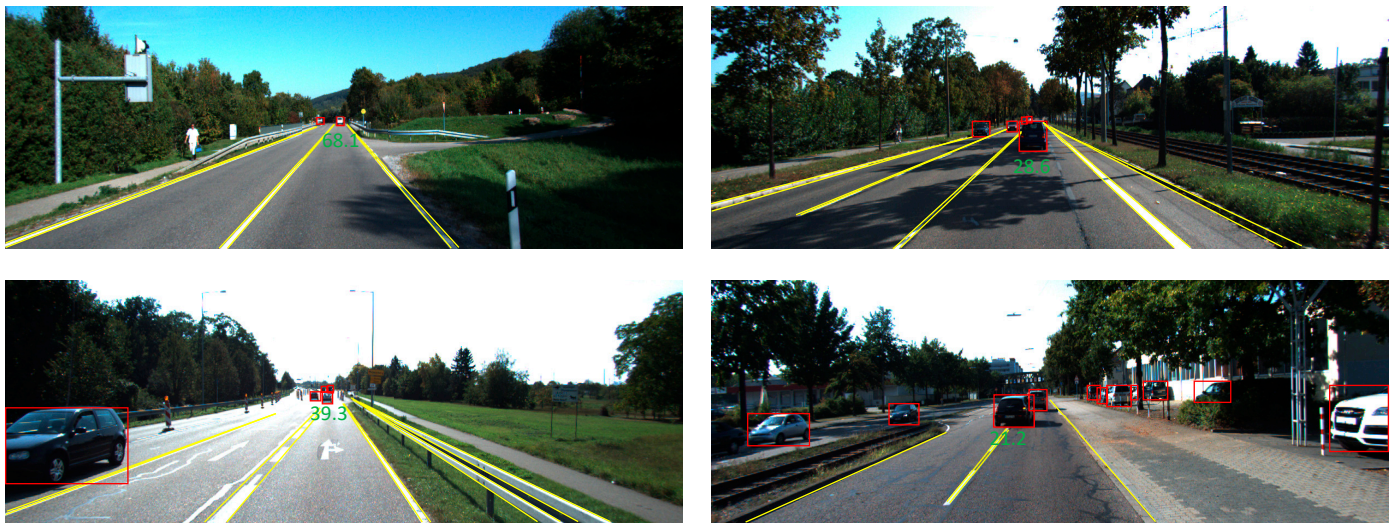


Figure 9. Visualization of ranging results. (The red box represents the vehicle target, and the green number represents the distance of the nearest vehicle.)

4.2. Dataset Validation

To demonstrate the wide applicability of the model, the ranging model proposed in this paper was validated on two datasets, KITTI and TuSimple. The KITTI 2012 dataset contains 7480 training images and 7517 test images, while the TuSimple dataset contains 3626 training images and 2782 test images. The validation results are shown in the Figure 10.

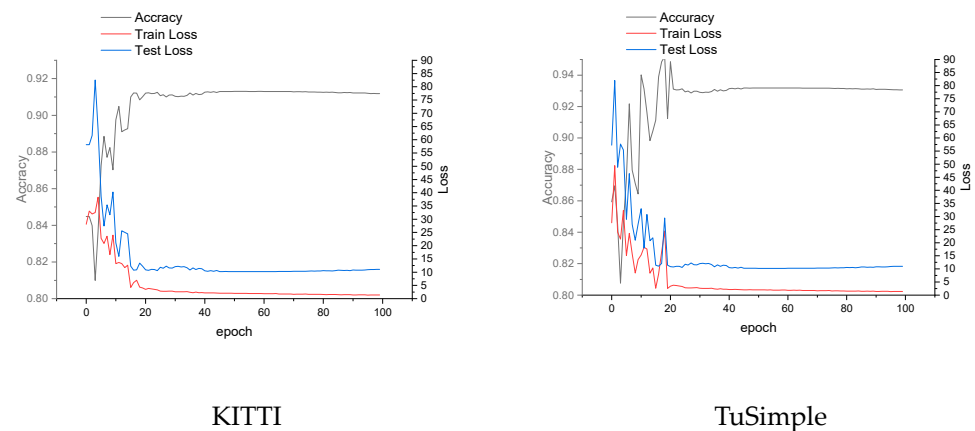


Figure 10. Validation results.

From Figure 10, it can be seen that the model in this article performs well on the KITTI dataset, with a relative error (AbsRel) of 8.3% after 100 network iterations. On the TuSimple dataset, the accuracy of distance measurement in this paper reached 93.1%, proving that the model can be widely used in various scenarios. The comparison between the model and other literature is shown in Table 3.

The model in this article has shown good performance in parameters such as AbsRel, RMS, and $\delta < 1.25^2$, which is an improvement in performance compared to previous models. During driving, vehicle suspension and road undulations can cause changes in camera posture, which in turn affects ranging accuracy. Compared with reference [11], the estimation of camera attitude effectively improves the ranging accuracy. Compared to reference [10], multiple geometric schemes as prior conditions for the network have better robustness to changes in road undulations and vehicle widths. However, the ranging networks also requires the consumption of more computing resources.

Table 3. Result comparison table based on KITTI.

	3Dbbox	Liu et al. [10]	Zhao et al. [11]	Proposed Method
AbsRel ↓	0.222	0.095	0.085	0.083
SqRel ↓	1.863	0.454	0.375	0.402
RMS ↓	7.696	4.728	4.114	4.100
RMSlog ↓	0.228	0.153	0.108	0.121
$\delta < 1.25^1$ ↑	0.659	0.903	0.974	0.952
$\delta < 1.25^2$ ↑	0.966	0.995	0.997	0.998
$\delta < 1.25^3$ ↑	0.994	1.000	1.000	1.000

↑ for higher value and better performance, ↓ for lower value and better performance.

4.3. Continuous Image Verification

The vehicle visual system experimental platform in this article consists of an experimental vehicle, one monocular onboard camera, one Jetson Nano (NVIDIA, Santa Clara, CA, USA), and a display, as shown in Figure 11. The monocular camera is installed on the inside of the front windshield of the experimental vehicle and connected to the Xavier development board through a USB conversion cable. The entire system is powered by the onboard inverter.



Figure 11. Experiment based on actual vehicles: (a) system experimental platform; (b) examples of dataset.

The experiment was conducted on a structured road, with the vehicle moving at a constant speed and the target vehicle stationary in the side lane. A total of 1415 frames of images were captured by the camera, and six frames were selected based on the markers placed every 10 m on the lane. The actual longitudinal distance was obtained through distance markers, and the ranging results of our method are shown in Table 4.

Table 4. Results based on actual vehicles.

No.	Pitch Angle			Yaw Angle			Distance		
	True Value	Estimated Value	Relative Error/%	True Value	Estimated Value	Relative Error/%	True Value/m	Estimated Value/m	Relative Error/%
1	1.31	1.86	41.98	2.78	3.02	8.63	27.3	26.94	1.32
2	3.27	3.49	6.73	5.08	5.24	3.15	37.3	36.7	1.61
3	4.96	5.08	2.42	7.81	8.21	5.12	47.3	46.98	2.79
4	7.53	7.97	5.84	9.36	9.69	3.53	57.3	55.77	2.67
5	11.87	12.35	4.04	14.85	15.42	3.84	67.3	65.04	3.36
6	14.38	14.89	3.55	16.97	17.51	3.18	77.3	74.51	3.61

This article proposes a forward vehicle monocular ranging method, which includes camera attitude estimation and a distance estimation network. Common vehicle ranging solutions often use a combination of cameras and radars [12]. This approach has better robustness to weather changes. Compared to radar, lidar, stereo-cameras, etc., monocular cameras are more economical and suitable for vehicles with limited budgets. In addition, the monocular camera has a compact structure and is easier to install in different positions of the vehicle. The data processing of monocular cameras is usually relatively simple and requires low computing resources.

Another monocular ranging scheme is to use a 3D detection frame [13]. This approach provides location information and size estimation of objects, which helps to understand their physical properties. However, it is more complex than two-dimensional detection and more sensitive to noise, especially when the appearance of the occluded object is similar to that of the occluded object.

5. Conclusions

This article proposes a forward vehicle monocular ranging method, which includes camera attitude angle estimation correction and a distance estimation network, and validates KITTI, TuSimple, and continuous scenes. Compared with previous algorithms, this method has improved performance. The model mainly has the following characteristics: (1) For the camera pose estimation scheme, it is achieved through road vanishing points. The specific process is edge feature point extraction, edge feature line segment generation, road vanishing point voting, and camera pose angle estimation. (2) For the distance measurement method of the fusion network of the grounding point and vehicle width, a distance measurement network with multiple image size levels and considering contour information was designed, effectively integrating the distance measurement method of the grounding point and vehicle width. (3) The proposed approach was verified on KITTI data with a relative error of 8.3%, showing good performance on parameters such as AbsRel, RMS, and $\delta < 1.25^2$. On the TuSimple dataset, the relative error was 6.9%, and tracking distance measurement can also be achieved in continuous driving scenarios, which has improved performance compared to previous algorithms.

Author Contributions: Conceptualization, D.X.; methodology, J.L.; software, D.X.; validation, J.L. and D.X.; data curation, J.L.; writing—original draft preparation, D.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by dual rotor wheeled motor driven vehicle variable voltage charging regenerative braking and hydraulic ABS coordinated control (51875258).

Data Availability Statement: The original contributions presented in the study are included in the article; further inquiries can be directed to the corresponding author.

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

References

1. Duan, X.T.; Zhou, Y.K.; Tian, D.X.; Zheng, K.X.; Zhou, J.S.; Shun, Y.F. A Review of the Application of Deep Learning in the Field of Autonomous Driving. *Unmanned Syst. Technol.* **2021**, *4*, 1–27.
2. Kim, S.H.; Hwang, Y. A survey on deep learning based methods and datasets for monocular 3D object detection. *Electronics* **2021**, *10*, 517. [\[CrossRef\]](#)
3. Chen, Y.C.; Su, T.F.; Lai, S.H. Integrated vehicle and lane detection with distance estimation. In Proceedings of the Computer Vision-ACCV 2014 Workshops, Singapore, 1–2 November 2014; pp. 473–485.
4. Seo, Y.W.; Rajkumar, R. Use of a monocular camera to analyze a ground vehicle's lateral movements for reliable autonomous city driving. In Proceedings of the IEEE IROS Workshop on Planning, Perception and Navigation for Intelligent Vehicles, Tokyo, Japan, May 2013; pp. 197–203.
5. Park, K.Y.; Hwang, S.Y. Robust range estimation with a monocular camera for vision-based forward collision warning system. *Sci. World J.* **2014**, *2014*, 923632. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Li, B.; Zhang, X.; Sato, M. Pitch angle estimation using a Vehicle-Mounted monocular camera for range measurement. In Proceedings of the 2014 12th International Conference on Signal Processing (ICSP), Hangzhou, China, 19–23 October 2014; pp. 1161–1168.

7. Stein, G.P.; Mano, O.; Shashua, A. Vision-based ACC with a single camera: Bounds on range and range rate accuracy. In Proceedings of the IEEE IV 2003 Intelligent Vehicles Symposium, Proceedings (Cat. No. 03TH8683), Columbus, OH, USA, 9–11 June 2003; pp. 120–125.
8. Zhang, R. Research on Monocular Vision Vehicle Fusion Ranging Algorithm in Dynamic Driving Scenarios. Master's thesis, Jiangsu University, Zhenjiang, China, 2020.
9. Liu, J.; Hou, S.H.; Zhang, K.; Yan, X.J. Vehicle distance measurement based on monocular vision vehicle attitude angle estimation and inverse perspective transformation. *J. Agric. Eng.* **2018**, *34*, 70–76.
10. The KITTI Vision Benchmark Suite. Available online: <https://www.cvlibs.net/datasets/kitti/> (accessed on 18 July 2024).
11. Zhao, H.X. Research on Monocular Ranging and Speed Measurement Algorithms for Driving Assistance Systems. Master's thesis, Changchun University, Jilin, China, 2022.
12. Wang, X.; Xu, L.; Sun, H.; Xin, J.; Zheng, N. On-Road Vehicle Detection and Tracking Using MMW Radar and Monovision Fusion. *IEEE Trans. Intell. Transp. Syst.* **2016**, *17*, 2075–2084. [[CrossRef](#)]
13. Zhe, T.; Huang, L.; Wu, Q.; Zhang, J.; Pei, C.; Li, L. Inter-Vehicle Distance Estimation Method Based on Monocular Vision Using 3D Detection. *IEEE Trans. Veh. Technol.* **2020**, *69*, 4907–4919. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.