

Review

A Lightweight Approach to Understand Forest Roads for New Energy Vehicles

Luping Wang^{1,*}, Yuan Feng¹, Shanshan Wang², and Hui Wei³

¹ Laboratory of 3D Scene Understanding and Visual Navigation, School of Mechanical Engineering, University of Shanghai for Science and Technology, No. 516 Jungong Road, Shanghai200093, China

² Intel Asia-Pacific Research & Development Ltd., No.880 Zixing Road, Shanghai201100, China

³ Laboratory of Algorithms for Cognitive Models, School of Computer Science, Fudan University, No. 825 Zhangheng Road, Shanghai201203, China

* Correspondence: 15110240007@fudan.edu.cn

Received: 16 June 2024; Revised: 21 October 2024; Accepted: 24 October 2024; Published: 11 November 2024

Abstract: Scene understanding is a core issue for autonomous vehicles. However, its implementation has been thwarted by various outstanding issues, such as understanding forest roads in unknown field environments. Traditional three-dimensional (3D) point clouds or 3D estimation of fused data consume large amounts of memory and energy, making these methods less reliable in new energy vehicles with limited computational, memory, and energy resources. In this study, we propose a lightweight method to understand forest roads using a low-cost monocular camera. We extracted and clustered spatially similar texture projections based on oblique effect. Through the relative relationship between vanishing points and texture projections, contour lines can be estimated. After that, searching for the corresponding supporting evidence lines, we can segment the surface of the forest road, which can provide a decision basis for the automatic driving control system of new energy vehicles with limited resources. Unlike deep learning methods that are extremely resource-consuming, the proposed method requires no prior training, no calibration, and no internal parameters of the camera. At the same time, pure geometric reasoning makes the method robust to the ever-changing colors and lighting in the forest environment. The percentage of correctly classified pixels is compared to the ground truth. The experimental results show that the method can successfully understand forest roads and meet the requirements of autonomous navigation in forest environments for new energy vehicles with limited resources.

Keywords: forest road; scene understanding; lightweight; autonomous driving; new energy vehicle

1. Introduction

Scene understanding is a key problem in autonomous driving. It has a multitude of unsolved problems, such as understanding forest roads in complex field environments. Understanding forest roads using a resource-constrained system with limited computation, memory, and energy resources remains a considerable challenge for new energy vehicle.

Monocular camera-based scene understanding systems have advantages in terms of energy consumption, computational efficiency, and cost control over 3D point cloud-based or multi-sensor data fusion approaches. Humans interpret semantic information of a scene through interpretable visual cues in the scene [1,2]. Although it is more difficult for humans to estimate the depth of isolated points in a scene, it is possible to utilize the relative geometric relationship cues in the scene to reason about the structure of the scene, which plays a crucial role in understanding the scene.

Unlike indoor and urban scenes, forest roads are always disturbed by soil, vegetation and weeds. As shown in the Figure 1, the rugged road surfaces are subjected to different lighting, diverse colors and



unpredictable vegetation, making most of the edges and surfaces of the forest roads appear fragmented. Nonetheless, their spatial texture (2D projections with different configurations) maintains traceable geometric features in terms of location and orientation, providing a basis for understanding forest roads.



Figure 1. An unstructured forest road disturbed by soil, trees and weeds.

In this study, we present a method for understanding forest roads in a field environment without prior training, using a low-cost monocular camera. First, based on the oblique effect, we extracted clusters of tree texture projections. Then, by analyzing them in relation to vanishing points, the contour lines of forest roads were approximated. Finally, the forest road surfaces can be estimated based on the position and orientation of the corresponding evidence lines.

For the current popular deep learning method, it requires large-scale labeled data for training, and is extremely dependent on a large number of computational and storage resources, such as high-performance graphics processors (GPUs) and large-capacity storage devices and other hardware devices. These consume a lot of energy and are difficult to apply to new energy vehicles with energy constraints. Therefore, unlike machine learning methods, the method presented in this paper does not require prior training and has lower hardware requirements, which has a lower energy consumption. Interpretable geometric inference makes the proposed method robust to unexpected colors, lighting, and materials in forest environments. Understanding forest roads through monocular cameras without precise depth or knowledge of the camera's intrinsic parameters makes it more practical and affordable.

The percentage of correctly categorized pixels was evaluated by comparing the estimated forest roads with the ground truth. The experimental results demonstrate that the proposed geometric inference-based method is able to understand forest roads, and our method is energy-saving, efficient, lightweight, and has a promising application in the visual navigation of new energy vehicles.

2. Related Work

There are some classical approaches for scene reconstruction such as SFM [3] and visual SLAM [4]. Researchers have proposed a 3D point cloud based tracking model to solve the occlusion problem in large clusters of featureless objects [5]. However, the acquisition and processing of 3D point cloud data is more complex, computationally intensive, and energy consuming, and appears to be impotent when interpreting geometric structures such as edges, textures, and spatial planes.

A great deal of research has been done to understand structured environments such as indoor or urban scenes. Significant progress has been made in areas such as indoor spatial layout and scene details through monocular cameras [6,7]. Research has also been done to propose a framework for detecting and segmenting puddles on roads in unstructured environments from dashboard camera images [8]. Some geometric models were developed to detect walkable surfaces, but these models were not validated in field environments [9–11]. Real-time pixel semantic classification and segmentation using hyperspectral images is proposed to extract, filter and approximate polygonal

objects [12]. The problem of spatial layout in outdoor environments is solved by projecting at right angles to space [13,14]. However, these scene-oriented structured methods cannot work in field road environments due to their limited interpretation of natural features.

For unstructured road environments, data-driven algorithms have been adopted and extended for object recognition and semantic segmentation [15–17]. For semantic scene segmentation in unstructured environments, the researcher proposes a lightweight segmentation model improved in the DeepLabv3+ framework [18]. In addition, the researchers proposed a terrain classification framework for pixel semantic segmentation of images using an unsupervised proprioceptive classifier learned from vehicle-terrain interaction sounds [19]. However, these data-driven models do not take into account unstructured road surfaces in forest scenes well.

Some scholars have studied obstacle avoidance and traversable regions for autonomous navigation. Based on the shape and attitude of obstacles, a pure geometric model was developed to avoid obstacles [20]. Traversability maps for unstructured environments were obtained and the optimal paths were estimated based on the specific characteristics of the robot [21, 22]. Another approach uses irregular angle projections to model unstructured paths in field environments but cannot handle forest road without angle projections [23,24]. Most of these methods focus only on categorizing traversable or non-traversable areas and therefore cannot understand forest roads.

For natural scenes, a method for detecting vanishing points is proposed that combines contour-based vanishing point edges [25]. Some researchers have also proposed a dense semantic segmentation algorithm based on resolution pyramid and heterogeneous feature fusion [26]. There have been attempts to perform semantic segmentation in unstructured scenes by stereo visual odometry [27]. In order to cope with changes in the environmental terrain, an ensemble category semantic segmentation method has been proposed to obtain different categories, such as sky and obstacles in off-road environments [28]. However, these methods lack edge resolution and are difficult to apply to forest roads.

Since most methods are designed for scene understanding of structured paths and require prior training, it is necessary to propose an energy-saving, efficient, and lightweight method that is able to understand unstructured paths in forest environments by means of pure geometric modeling without the need for prior training and external equipment, so as to satisfy the requirements of resource-constrained new energy vehicles to achieve autonomous navigation in field environments.

3. Understanding a Forest Road

In the case of forest roads, the surface, although always fragmented, still has a particular spatial texture. These textures are projected as two-dimensional projections with different shapes that can be considered as visual cues. By analyzing the position and orientation of these texture projections, their spatial morphology can be inferred, thus contributing to the understanding of such forest roads.

3.1. Preprocessing

First, lines and vanishing points (VPs) are detected [29,30] as follows:

$$line_i = [p_1^i, p_2^i] = [x_1^i, y_1^i, x_2^i, y_2^i]; i \in [1, \mathbb{N}], \tag{1}$$

$$VP = [P_1; P_2; P_3;] \tag{2}$$

where \mathbb{N} is the number of lines. P_1, P_2 and P_3 are layout vanishing points.

3.2. Tree Trunk Texture Projection

Inspired by the oblique effect, neurons prefer to respond to horizontal and vertical stimuli more than to diagonal. Therefore, trees and plant bodies, which tend to grow vertically or horizontally, help estimate road layouts in field forest environments. A pair of tree bodies texture projection can be represented by a composition of two lines with geometric constraints, which can be defined as follows:

$$\mathbb{T} = \{line_w, line_e\}; w, e \in [1, \mathbb{N}] \tag{3}$$

$$d_{w,e} = \text{distance}(p_m^w, p_m^e) \tag{4}$$

$$\text{St. min } d_{w,e}; \min\Theta(\text{line}_w, \text{line}_e) \tag{5}$$

Here \mathbb{T} is a pair of tree body projection that includes two lines line_w and line_e . Here, $d_{w,e}$ is the Euclidean distance between two midpoints ($p_m^w = [x_m^w, y_m^w]$ and $p_m^e = [x_m^e, y_m^e]$) of line_w and line_e , respectively. A smaller distance means that the two lines are closer together, indicating better integrity of the two lines. Θ is a function to compute the angle of these two lines. Smaller angle values indicate that the two lines are more similar in orientation. The smaller the distance and angle values of the two lines indicate that this combination is more likely to be a two-dimensional projection of the spatial trunk texture.

A tree trunk texture projection has two lines. Through corresponding positions and orientations, it is possible to assign them to different clusters. Assuming that $\mathbb{Y}(\mathbb{T})$ is a function to cluster tree trunk texture projection:

$$\mathbb{Y}(\mathbb{T}) = \begin{cases} 1, & \min(x_m^w, x_m^e) x_{P_1}, \beta_w \rightarrow \pi/2, \beta_e \rightarrow \pi/2 \\ 2, & \max(x_m^w, x_m^e) x_{P_1}, \beta_w \rightarrow \pi/2, \beta_e \rightarrow \pi/2 \end{cases} \tag{6}$$

It indicates that a tree trunk texture projection can be assigned to different clusters, which are written as $\mathbb{Y}^k, k \in 1, 2$. Accordingly, corresponding clusters can be extracted, as shown in Figure 2.

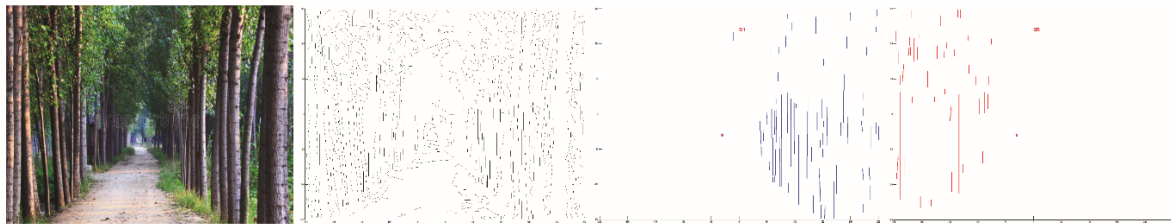


Figure 2. Preprocessing and tree texture projections. Left: input image. Second: lines. Third and Fourth: clusters of \mathbb{Y}^1 and \mathbb{Y}^2 , respectively.

3.3. Contour Layout Lines

Because a forest road is always covered by unexpected grass, vegetation or soil, its shape appears fragmentary and irregular. Therefore, it is necessary to estimate contour layout lines that can coarsely infer the road.

Given a classified $\mathbb{Y}^k = [\text{line}_w^k, \text{line}_e^k]$, $p_b^{k,w}$ and $p_b^{k,e}$ means the bottom points in corresponding line_w^k and line_e^k , respectively. The following can be modelled:

$$\eta_b^{k,w} = \Gamma([P_1, p_b^{k,w}], [\sigma_{min}, \sigma_{max}]) \tag{7}$$

$$\eta_b^{k,e} = \Gamma([P_1, p_b^{k,e}], [\sigma_{min}, \sigma_{max}]) \tag{8}$$

Here Γ is a function to find the point of intersection of two lines. $\sigma_{min} = [-W/2, -H/2]$, and $\sigma_{max} = [W/2, H/2]$, in which W and H are the height and width of a monocular capture. $\eta_b^{k,w}$ means an intersection point of two lines, which are represented by $[P_1, p_b^{k,w}]$ and $[\sigma_{min}, \sigma_{max}]$.

For all lines in cluster $k = 1$, the corresponding points $\eta_b^{k,w}$ and $\eta_b^{k,e}$ are computed, as shown in Figure 3 (red points in left and middle images). Then these point that has the minimum x value is determined and written as $\eta_b^{*k=1}$. This point and P_1 can shape a contour line. Similarly, for cluster $k = 2$, the point that has the maximal x value is described as $\eta_b^{*k=2}$, which can shape another contour line. Accordingly, two virtual lines can be modelled as following:

$$\mathbb{U} = \{\Psi^1, \Psi^2\}; \tag{9}$$

$$\Psi^1 = [\eta_b^{*k=1}, P_1] \tag{10}$$

$$\Psi^2 = [\eta_b^{*k=2}, P_1] \tag{11}$$

Here \cup represents the layout lines of a forest road. $\eta_b^{*k=1}$ represents the point has the minimum x value in $\eta_b^{k=1,w}$ and $\eta_b^{k=1,e}$. Similarly, $\eta_b^{*k=2}$ represents the point has the minimum x value in $\eta_b^{k=2,w}$ and $\eta_b^{k=2,e}$. These two lines Ψ^1 and Ψ^2 can be seen as two contour lines of the road, as shown in Figure 3 (pink lines).

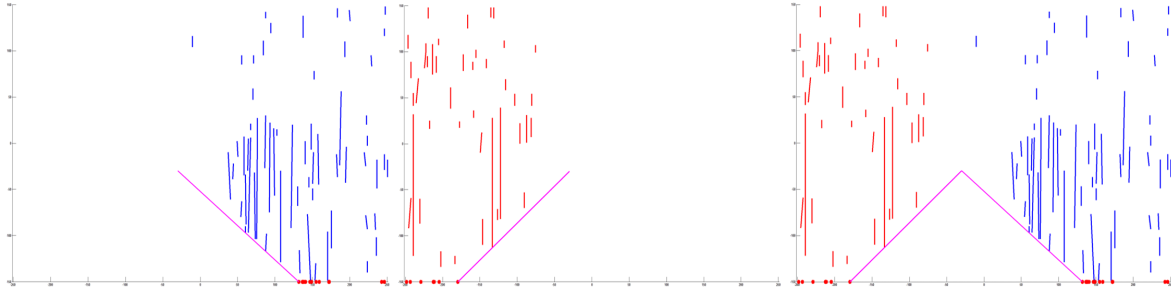


Figure 3. Contour layout lines. Left: Ψ^1 . Middle: Ψ^2 . Right: layout lines. Contour layout lines of a forest road can be described as two virtual lines Ψ^1 and Ψ^2 .

3.4. Evidence Surfaces

For Ψ^1 , in order to determine whether there are other lines following these two lines, for $line_j (j \in \mathbb{N})$, the condition can be founded:

$$\mathbb{E} = \{line_j\}; j \in \mathbb{N} \tag{12}$$

$$s.t. MIN(\delta_j^1); MIN(\delta_j^2); \tag{13}$$

$$\delta_j^1 = \|\Theta(line_j, \Psi^1)\| + \|p_m^j - \Psi^1\| \tag{14}$$

$$\delta_j^2 = \|\Theta(line_j, \Psi^2)\| + \|p_m^j - \Psi^2\| \tag{15}$$

Here, \mathbb{E} is a set of evidence lines for layout lines. δ_j^1 and δ_j^2 are relative relationships between $line_j$, Ψ^1 and Ψ^2 , respectively. Θ is a function computing angle of two lines, and the angle representing orientation constraint. Smaller value of the angle means that $line_j$ is more likely to have similar orientation to Ψ^1 . p_m^j is midpoint of $line_j$. $\|p_m^j - \Psi^1\|$ is the distance from p_m^j to Ψ^1 . Smaller value of the distance means that $line_j$ is closer to Ψ^1 . δ_j^1 means the normalized value. Hence, $line_j$ follows the above geometric constraints, which indicates that the line can be considered evidence supporting the contour line Ψ^1 . Similarly, evidence lines for Ψ^2 can be extracted, as shown in Figure 4.

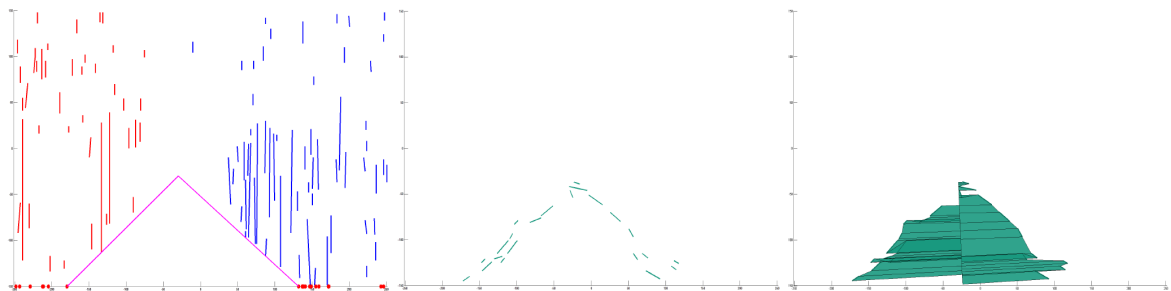


Figure 4. Evidence surfaces. Left: layout lines. Middle: evidence lines. Right: surfaces. Through geometric constraints, the evidence lines supporting the contour layout lines are extracted, and corresponding surfaces are approximately reshaped as a forest road.

The corresponding surfaces can be approximated by irregular, fragmented lines of evidence. Assuming that $line_j$ is an evidence line in \mathbb{E} , then the following can be modelled:

$$S^j = \left[\Gamma([p_m^n, P_1], [p_1^j, P_2]); \Gamma([p_2^j, P_2], [p_m^n, P_1]); p_2^j; p_1^j; \right]^T, j \in \mathbb{E} \tag{16}$$

Here S^j represents a surface of an evidence line in \mathbb{E} . p_m^n is the midpoint of virtual line $[\eta_b^{*k=1}, \eta_b^{*k=2}]$. For all evidence lines in \mathbb{E} , their corresponding road surfaces can be shaped as shown in Figure 4.

4. Experimental Results

4.1. Requirement

Our research is based entirely on pure geometric reasoning, using only a low-cost monocular camera (RGB image) to understand various unstructured paths in a forest environment. Since no precise depth data or any prior training is required, our experiments were conducted on a computer with an Intel Core i7-6500 2.50 GHz CPU and 2GB of RAM, with no additional high-performance graphics processor (H-GPU), no calibration, and no internal parameters of the camera (IPC), as shown in Table 1. Obviously, the lightweight method utilizing only one monocular camera has lower energy consumption and is more suitable for new energy vehicles with energy constraints.

Table 1. Requirement list.

Condition	RGB	Depth	Training	H-GPU	Calibration	IPC
Requirement	√	×	×	×	×	×

4.2. Evaluation

There are many datasets for understanding road scenes, for example, the widely used urban driving datasets (e.g., Cityscapes [31], KITTI [32]), which are structured road scenes; there are also road datasets for outdoor environments (e.g., TAS500 [33]), but few of them contain forest roads. Our approach aims to understand unstructured roads in forested environments with a low-cost monocular camera; therefore, we build a new dataset with 253 different Unstructured Forest Road (UFR) images, which help us evaluate the effectiveness of the proposed method.

A forest road is shown in Figure 5. By comparing the estimated forest road surfaces to the ground truth, the percentage of correctly classified pixels was evaluated, as shown in Table 2. It calculates the correctly classified pixels, i.e., $IoU = \frac{TP}{TP + FP + FN}$, where TP , FP , and FN are the numbers of true positive, false positive, and false negative pixels, respectively. The presented method can not only understand structured roads, but also can account for fragmentary road surfaces in forest environments. The results demonstrate that our algorithm has advantages in understanding such forest roads in field environments.

Table 2. Error of understanding unstructured forest roads.

Method	IoU
H.W. [2]	33.2%
Wang [10]	52.3%
Our method	78.7%

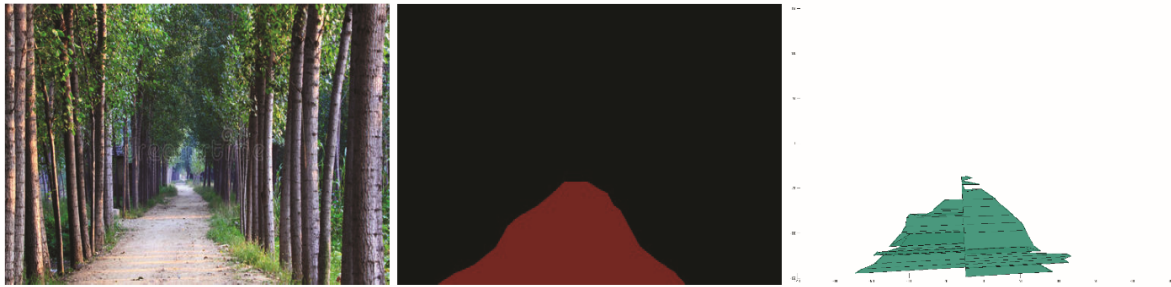


Figure 5. Understanding a forest road. Left: input image. Second: ground truth. Third: understanding. Compared to the ground truth, our approach can understand rough unstructured surfaces of a forest road without any prior training.

4.3. Comparison

Comparisons between the previous algorithms and the proposed method were made on the UFR dataset, with the results shown in Figure 6. Previous approaches approximate a scene using only spatial rectangles [2] and spatial right angles [10], but there is few complete shape of rectangles or right angles in forest environments. By contrast, our algorithm, based on assumption of contour line and evidences, has the advantages of accounting for fragmentary surfaces of forest roads that are always covered by unexpected water and side structures such as vegetation, soil, and snow in forest road environments.

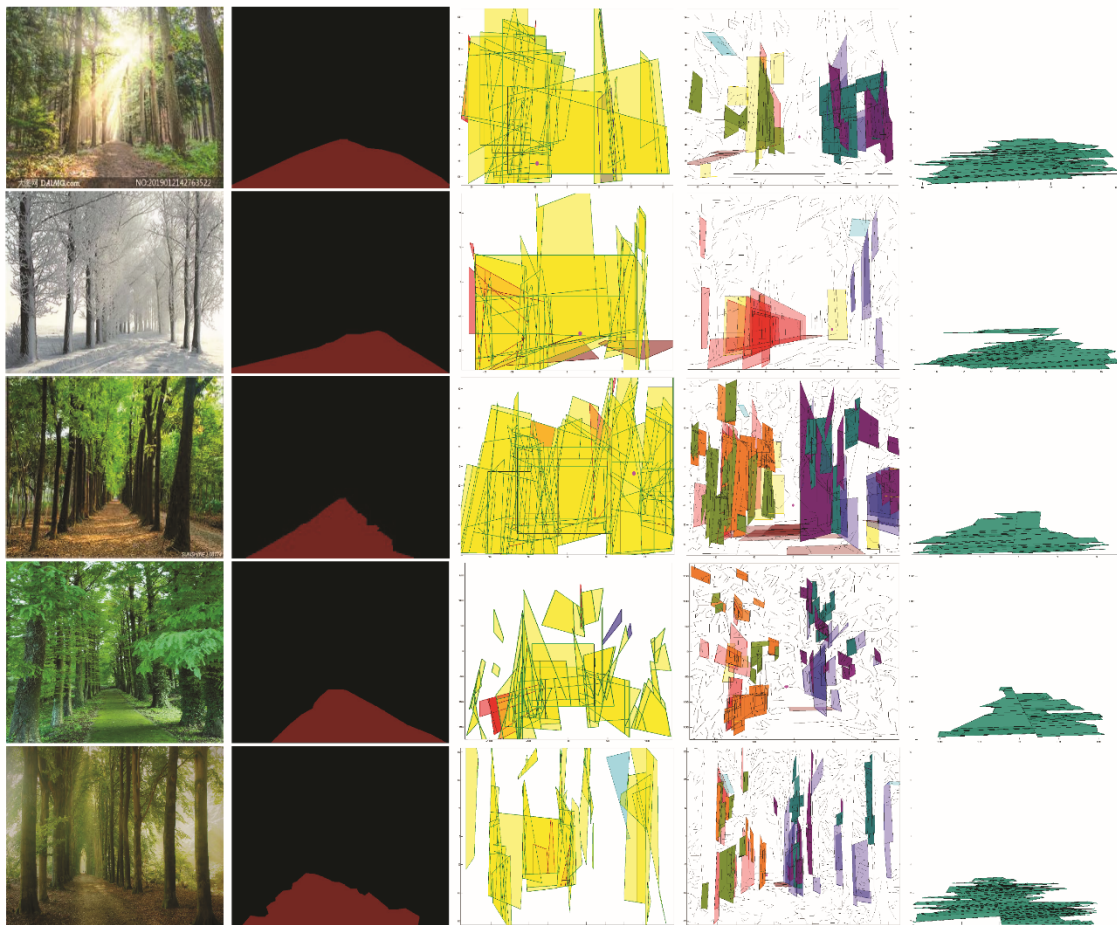


Figure 6. Experimental comparisons on UFR dataset. First column: forest roads. Second column: ground truth. Third column: H.W.'s method [2]. Fourth column: Wang's method [10]. Right column: our understanding. In contrast, our method can understand forest roads covered by water and side covers, e.g., vegetation, soil, and snow.

Furthermore, more experiments were performed on forest roads within diverse environments of different color and illumination. Since the method adopts geometric inferences, as shown in Figure 7, the approach can successfully understand forest roads, which is robust against changes in illumination and color.

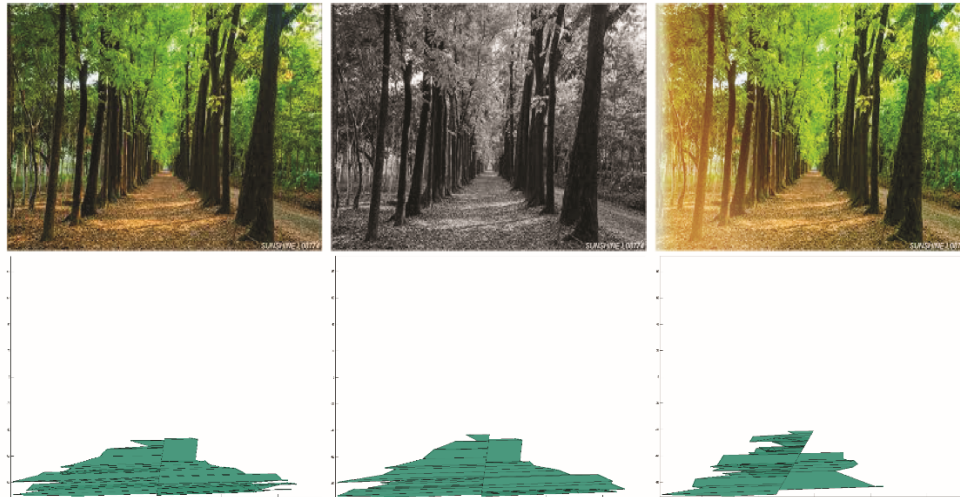


Figure 7. Robustness experiment. First column: original environment. From second column to right column: changes in color and illumination, respectively. The proposed method is robust against changes in color and illumination.

Furthermore, more experiments were performed on best and worst cases, as shown in Figure 8. Since the method relies on scene line texture features, the results are worst when there are fewer lines and best when there are sufficient texture features.

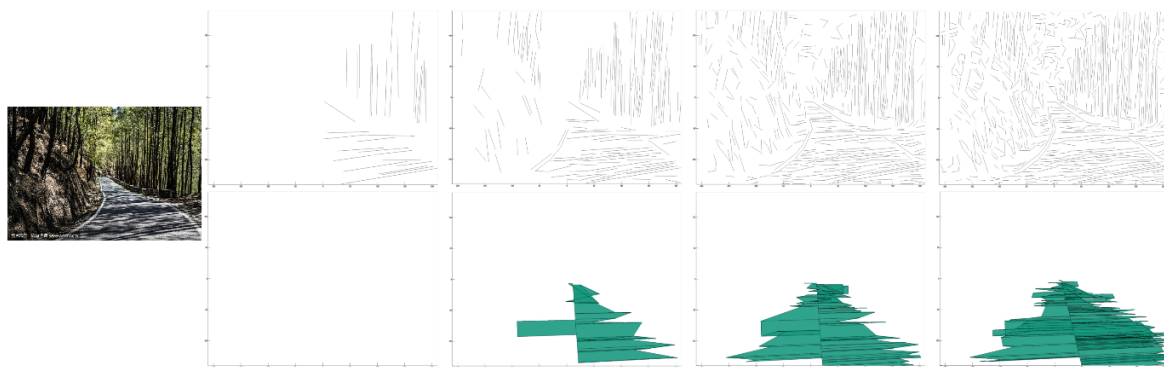


Figure 8. Best and worst cases. First row: different detected lines from images. Second row: best and worst cases. Since the method relies on line extraction, the algorithm may not be able to help if there is a lack of corresponding texture features in the scene.

More experiments were compared with deep learning based approaches and the results are shown in Figure 9. The framework is end-to-end and has difficulty interpreting geometric structural features of field roads. In contrast, our algorithm can interpret geometric cues and understand field roads without prior training.

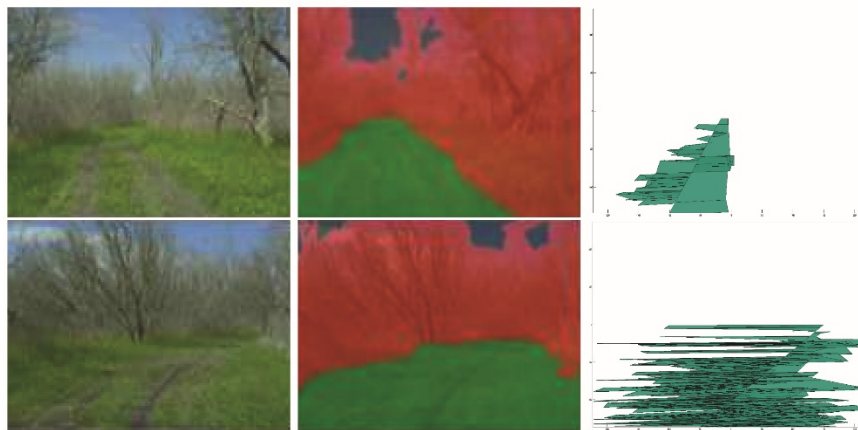


Figure 9. Experimental comparisons to deep learning methods. First column: forest ways. Second column: Duraisamy's method [34]. Third column: our understanding. Our approach can understand forest ways without prior training.

4.4. Discussion

Experimental results show that traditional methods rely on structured features of scene roads, which makes them unsuitable for functioning in wild forest road environments full of unstructured disturbances. Deep learning methods are unable to handle complex and variable unstructured scene information (e. g., layout, location, and orientation), and are often overly complex and require high-performance GPUs, which are costly to implement in terms of energy consumption. Compared with traditional methods, the method proposed in this paper has advantages in dealing with unstructured environments without prior training, has lower energy consumption, and meets the needs of energy-constrained new energy vehicles to navigate autonomously in forest environments.

Previous methods have only used spatial rectangles and spatial right angles to approximate scenes, but forest environments rarely have complete rectangular or right-angle shapes. Inspired by the oblique effect, neurons preferred to respond to stimuli orientated horizontally and vertically rather than to angles. Thus, trees and plant bodies that tend to grow vertically or horizontally help to estimate road layouts in wild forest environments. In contrast, our algorithm, which is based on the forest vertical trunk hypothesis and the evidence hypothesis, has the advantage of utilizing trunk features on both sides of the forest road environment.

5. Conclusion

The current work proposes an energy-saving, efficient, and lightweight forest road understanding method to help new energy vehicles with limited resources to understand unstructured forest roads using a low-cost, low-power monocular camera without the need for high power consumption, high-performance computing, and any prior training. First, forest road scene edges and lines are extracted. Then, the initial contour of the forest road is estimated by the relationship between the trunk texture projection and the spatial unstructured extinction point. After searching a large number of scattered evidence lines, the forest road plane can be estimated. Unlike the data-driven method, the proposed method does not require prior training and does not need a high-performance GPU, which makes it more suitable for new energy vehicles with limited resources. Geometric reasoning also makes the proposed method robust to changes in light and color, making it more suitable for forest road environments. The estimated forest roads are compared with the true values and the percentage of correctly categorized pixels is measured. The experimental results show that the proposed method can effectively understand unstructured forest roads with low power consumption and low complexity, which has a promising application in new energy vehicle systems.

Author Contributions: L.W.: conceptualization, methodology, software, validation, writing—reviewing and editing; Y.F.: data curation, validation, writing—original draft preparation; S.W.: visualization, investigation; H.W.: supervision. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the NSFC Project (Project Nos. 62003212 and 61771146).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data will be made available on reasonable request.

Conflicts of Interest: The authors declare no competing interests.

References

1. Wei H.; Wang L. Visual navigation using projection of spatial right-angle in indoor environment. *IEEE Trans. Image Process.* **2018**, *27*, 3164–3177.
2. Wei H.; Wang L. Understanding of indoor scenes based on projection of spatial rectangles. *Pattern Recognit.* **2018**, *81*, 497–514.
3. Magerand L.; Del Bue A. Revisiting projective structure from motion: A robust and efficient incremental solution. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 430–443.
4. Besco's B.; Cadena C.; Neira J. Empty cities: A dynamic-object-invariant space for visual SLAM. *IEEE Trans. Robot.* **2021**, *37*, 433–451.
5. Cavagna A.; Melillo S.; Parisi L.; Ricci-Tersenghi F. Sparta tracking across occlusions via partitioning of 3d clouds of points. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *43*, 1394–1403.
6. Wang L.; Wei H. Reconstruction for indoor scenes based on an interpretable inference. *IEEE Trans. Artif. Intell.* **2021**, *2*, 251–259.
7. Wang L.; Wei H. Indoor scene understanding based on manhattan and non-manhattan projection of spatial right-angles. *J. Vis. Commun. Image Represent.* **2021**, *80*, 103307.
8. Kumar, A.; Choudhary, A. Water-puddle segmentation using deep learning in unstructured environments. In Proceedings of the 2023 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI), Singapore, 11–13 December 2023; pp. 1–6.
9. Wang L.; Wei H. Understanding of wheelchair ramp scenes for disabled people with visual impairments. *Eng. Appl. Artif. Intell.* **2020**, *90*, 103569.
10. Wang L.; Wei H. Curved alleyway understanding based on monocular vision in street scenes. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 8544–8563.
11. Wang L.; Wei H. Recognizing slanted deck scenes by non-manhattan spatial right angle projection. *IEEE Intell. Syst.* **2022**, *37*, 75–85.
12. Medellin, A.; Bhamri, A.; Langari, R.; Gopalswamy, S. Real-time semantic segmentation using hyperspectral images for unstructured and unknown environments. In Proceedings of the 13th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS), Athens, Greece, 31 October–2 November 2023; pp. 1–5.
13. Wang L.; Wei H. Understanding of curved corridor scenes based on projection of spatial right-angles. *IEEE Trans. Image Process.* **2020**, *29*, 9345–9359.
14. Wang L.; Wei H.; Hao Y. Vulnerable underground entrance understanding for visual surveillance systems. *Int. J. Crit. Infrastruct. Prot.* **2023**, *41*, 100589.
15. Nikolovski, G.; Reke, M.; Elsen, I.; Schiffer, S. Machine learning based 3d object detection for navigation in unstructured environments. In Proceedings of the IEEE Intelligent Vehicles Symposium Workshops (IV Workshops), Nagoya, Japan, 11–17 July 2021; pp. 236–242.
16. Wigness, M.; Rogers, J.G. Unsupervised semantic scene labeling for streaming data. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 5910–5919.
17. Humblot-Renaux G.; Marchegiani L.; Moeslund T.B.; Gade R. Navigation-oriented scene understanding for robotic autonomy: Learning to segment driveability in egocentric images. *IEEE Robot. Autom. Lett.* **2022**, *7*, 2913–2920.
18. Baheti B.; Innani S.; Gajre S.S.; Talbar S.N. Semantic scene segmentation in unstructured environment with modified deeplabv3+. *Pattern Recognit. Lett.* **2020**, *138*, 223–229.
19. Zurn J.; Burgard W.; Valada A. Self-supervised visual terrain classification from unsupervised acoustic feature learning. *IEEE Trans. Robot.* **2021**, *37*, 466–481.
20. Wang L.; Wei H. Avoiding non-manhattan obstacles based on projection of spatial corners in indoor environment. *IEEE/CAA J. Autom. Sin.* **2020**, *7*, 1190–1200.
21. Arena, P.; Blanco, C.F.; Noce, A.L.; Taffara, S.; Patane, L. Learning traversability map of different robotic platforms for unstructured terrains path planning. In Proceedings of the International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 19–24 July 2020; pp. 1–8.
22. Arena, P.; Pietro, F.D.; Noce, A.L.; Taffara, S.; Patane, L. Assessment of navigation capabilities of mini cheetah robot for monitoring of landslide terrains. In Proceedings of the 6th IEEE International Forum on Research and Technology for Society and Industry, RTSI 2021, Naples, Italy, 6–9 September 2021; pp. 540–545.
23. Wang L.; Wei H. Winding pathway understanding based on angle projections in a field environment. *Appl. Intell.* **2023**, *53*, 16859–16874.
24. Wang L.; Wei H. Bending path understanding based on angle projections in field environments. *J. Artif. Intell. Soft Comput. Res.* **2024**, *14*, 25–43.

25. Kloukiniotis, A.; Moustakas, K. Vanishing point detection based on the fusion of lidar and image data. In Proceedings of the 30th Mediterranean Conference on Control and Automation, MED2022, Vouliagmeni, Greece, 28 June–1 July 2022; pp. 688–692.
26. Orsic M.; Segvic S. Efficient semantic segmentation with pyramidal fusion. *Pattern Recognit.* **2021**, *110*, 107611.
27. Holder C.J.; Breckon T.P. Learning to drive: End-to-end off-road path prediction. *IEEE Intell. Transp. Syst. Mag.* **2021**, *13*, 217–221.
28. Viswanath, K.; Singh, K.; Jiang, P.; Sujit, P.B.; Saripalli, S. OFFSEG: A semantic segmentation framework for off-road driving. In Proceedings of the 17th IEEE International Conference on Automation Science and Engineering, CASE, Lyon, France, 23–27 August 2021; pp. 354–359.
29. Arbelaez, P.; Maire, M.; Fowlkes, C.; Malik, J. From contours to regions: An empirical evaluation. In Proceedings of the 17th 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 2294–2301.
30. Wang L.; Hao Y.; Wang S.; Wei H. Vanishing point estimation inspired by oblique effect in a field environment. *Cogn. Neurodynamics* **2024**, 1–16.
31. Cordts M.; Omran M.; Ramos S.; Rehfeld T.; Enzweiler M.; Benenson R.; Franke U.; Roth S.; Schiele B. The cityscapes dataset for semantic urban scene understanding. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 3213–3223.
32. Geiger A.; Lenz P.; Stiller C.; Urtasun R. Vision meets robotics: The KITTI dataset. *Int. J. Robotics Res.* **2013**, *32*, 1231–1237.
33. Metzger, A.; Mortimer, P.; Wuensche, H. A fine-grained dataset and its efficient semantic segmentation for unstructured driving scenarios. In Proceedings of the 25th International Conference on Pattern Recognition, ICPR, Milan, Italy, 10–15 January 2020; pp. 7892–7899.
34. Duraisamy, P.; Natarajan, S. Multi-sensor fusion based off-road drivable region detection and its ros implementation. In Proceedings of the 2023 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET), Chennai, India, 29–31 March 2023; pp. 1–5.