

Article

Towards Accurate Ground Plane Normal Estimation from Ego-Motion

Jiaxin Zhang ¹, Wei Sui ¹, Qian Zhang ¹, Tao Chen ² and Cong Yang ^{2,*}¹ Horizon Robotics, No. 9, FengHao East Road, Beijing 100094, China² School of Future Science and Engineering, Soochow University, Suzhou 215222, China

* Correspondence: cong.yang@suda.edu.cn

Abstract: In this paper, we introduce a novel approach for ground plane normal estimation of wheeled vehicles. In practice, the ground plane is dynamically changed due to braking and unstable road surface. As a result, the vehicle pose, especially the pitch angle, is oscillating from subtle to obvious. Thus, estimating ground plane normal is meaningful since it can be encoded to improve the robustness of various autonomous driving tasks (e.g., 3D object detection, road surface reconstruction, and trajectory planning). Our proposed method only uses odometry as input and estimates accurate ground plane normal vectors in real time. Particularly, it fully utilizes the underlying connection between the ego pose odometry (ego-motion) and its nearby ground plane. Built on that, an Invariant Extended Kalman Filter (IEKF) is designed to estimate the normal vector in the sensor's coordinate. Thus, our proposed method is simple yet efficient and supports both camera- and inertial-based odometry algorithms. Its usability and the marked improvement of robustness are validated through multiple experiments on public datasets. For instance, we achieve state-of-the-art accuracy on KITTI dataset with the estimated vector error of 0.39°.

Keywords: ground plane normal; autonomous driving; odometry; kalman filter



Citation: Zhang, J.; Sui, W.; Zhang, Q.; Chen, T.; Yang, C. Towards Accurate Ground Plane Normal Estimation from Ego-Motion. *Sensors* **2022**, *22*, 9375. <https://doi.org/10.3390/s22239375>

Academic Editors: Yi Zhang, Yan Yan and Li Yuan

Received: 29 September 2022

Accepted: 29 November 2022

Published: 1 December 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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

Accurate ground plane normal estimation is crucial to autonomous driving applications' perception, navigation, and planning. This is because the ground plane in the vehicle's coordinate is dynamically changed due to braking and unstable road surface (see Figure 1). As a result, the vehicle pose, especially the pitch angle, oscillates from subtle to obvious [1]. To improve the robustness of autonomous driving system, ground plane normal is estimated and encoded in vision-related tasks, including 3D object tracking [2], lane detection [3–7] and road segmentation [8–11], etc. For instance, the ground plane parameters are used for multi-camera calibration in many applications [12–14]. They are also employed to estimate the depth information of an object on the ground [15–17], and provide vital absolute scale information to the system [18]. In addition to the aforementioned tasks, existing image-based mapping [19] and Bird's-Eye-View (BEV) perception [20–22] algorithms are also sensitive to the accuracy of the ground plane normal parameters. For instance, some BEV-based algorithms are applied with inverse perspective mapping (IPM) with extrinsic parameters from the image plane to the ground plane, thereby mapping pixels from image space to BEV space.

However, estimating accurate ground plane normal in real-time is challenging, especially in a monocular setup. The main reason is that the subtle dynamics of the ground plane normal reflect little in image spaces. Traditional methods usually first estimate homography transform, then decompose it into ground plane normal and ego-motion [23,24]. Recently, some neural networks were proposed to estimate the depth and normal simultaneously at the pixel level, with photometric and geometric consistency [25–27]. However, these image-based methods suffer from inadequate accuracy due to a loose connection between the ground plane normal dynamics and image clues. Besides, most previous works simplify

(or assume) that the ground plane normal vector of a moving vehicle is constant, which is contrary to the facts. In practice, the normal vector is slightly oscillating when the vehicle moves, even if the road surface seems flat. For instance, a 4-wheel sedan moves along a straight street with a front-facing camera mounted on the top of the windshield. The camera pitch angle (relative to the ground) usually oscillates with an amplitude around 1° . Though such dynamics reflect little in image spaces, it can be easily observed in the BEV space after image projecting using IPM with fixed extrinsic (see Figure 2a and supplementary video for better visualization). This phenomenon is also positively verified by our quantitative and qualitative experiments in Section 5.

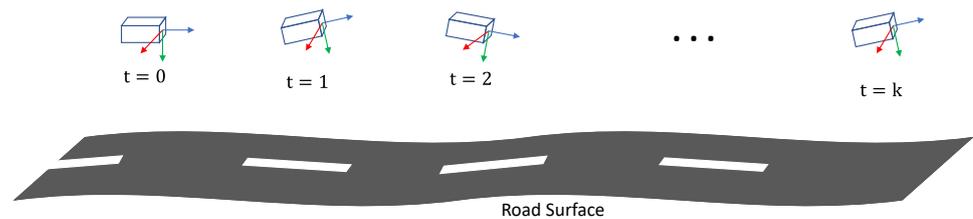


Figure 1. Illustration of a typical dynamic motion of a front-facing camera on a moving vehicle. The pitch angle (rotation around the x -axis) is actually oscillating with an amplitude of about 1° , though the vehicle moves straight and the road surface looks flat enough. Such pitch angle oscillation is amplified when the vehicle encounters imperfect road surfaces and speed bumps.

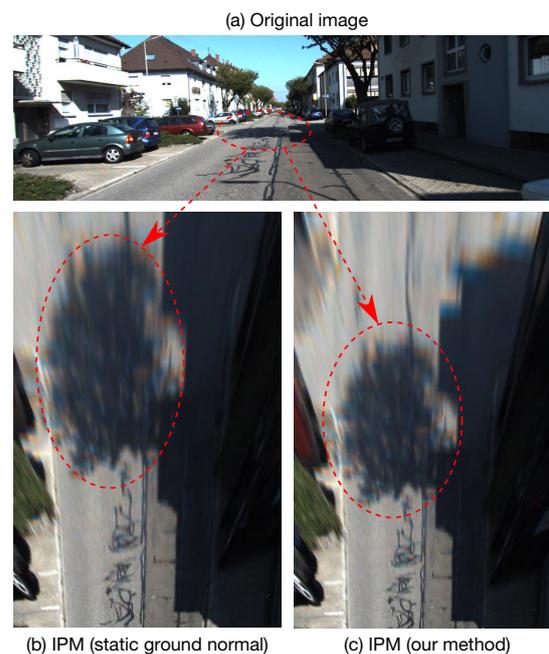


Figure 2. Comparison of IPM images before and after using our proposed method. (a) Original image from KITTI odometry dataset. (b) IPM image using fixed extrinsic from the camera to the ground. (c) IPM image using the dynamic extrinsic calculated by our proposed methods. It can be clearly observed that the image in (c) is more accurate. See our supplementary video for better visualization.

We introduce a simple yet efficient method to estimate ground plane normal from ego-motion. Particularly, our approach is compatible with ego-motion provided by SLAM (Simultaneous Localization And Mapping) and SfM (Structure from Motion) algorithms from various sensors (e.g., monocular camera and Inertial Measurement Unit (IMU)). To do so, we design an Invariant Extended Kalman Filter (IEKF) to model the dynamics of the vehicle's ego-motion and estimate the ground plane normal in real-time. Besides, our approach can be easily plugged into most autonomous driving systems that provide ego-

motion with little computational cost. As presented in Figure 2, after applying our proposed method, the image quality is dramatically improved. Our experiments in Section 5 verify its effectiveness: the estimated vector error is reduced from 3.02° with [26] to 0.39° with our proposed method on the KITTI dataset [28].

Succinctly, the main contributions of this work are as follows: (1) We introduce a simple yet efficient approach for real-time ground plane normal estimation. (2) The proposed method supports both camera- and inertial-based odometry algorithms thanks to the special design that fully utilizes the ego-motion information as input. Hopefully, our observations and contributions can encourage the community to develop more ground normal estimation methods towards robust autonomous driving systems in the real-world.

2. Related Works

We present a concise survey of existing ground normal estimation methods using depth sensors, stereo cameras, and monocular cameras. Some CNN-based methods are also discussed in this part. For a more detailed treatment of this topic in general, the recent compilation by Man and Weng [27] offers a sufficiently good review.

2.1. Ground Normal Estimation Using Depth Sensors

To obtain accurate ground plane parameters, using active depth sensors such as LiDAR and Time of Flight (ToF) is a reliable solution [29,30]. While the accurate 3D structure of the environments can be obtained in the form of point clouds (from LiDAR and ToF), ground plane parameters can be easily estimated by plane fitting. Built on that, the least square method is employed once the points belong to the ground [31,32]. For application, existing LiDAR-based works only are triggered to estimate ground planes in some challenging scenarios, such as offroad [33] and construction areas [34]. However, our proposed method takes ego-motion as input and can be easily plugged into most autonomous driving systems. As a result, our method is more general and can be employed in most driving scenes.

2.2. Ground Normal Estimation Using Stereo Cameras

Cheaper than active depth sensors such as LiDAR and ToF, stereo cameras are more accessible and can provide reasonable depth information through disparity. Similarly, most stereo-based methods are also designed to deal with particular cases, such as staircase [35] and cluttered urban environments [36]. However, they normally require good lighting conditions and rich textures. While depth and normals are highly related to 3D information, they are jointly trained with stereo images and consistency loss [37]. To directly model road surface with a plane normal, Stephen et al. estimate the ground plane based on disparity, thereby detecting and tracking obstacles and curbs [38]. Nikolay et al. also propose to use dense stereo disparity for ground plane normal estimation [39]. These disparity-based methods usually focus on analyzing the ground plane together with objects and the 3D structure of the road in detail. In comparison, our approach only requires a monocular camera or even IMU-only odometry to obtain high-accuracy ground plane normals in real time.

2.3. Ground Normal Estimation Using Monocular Camera

Ground plane estimation from a monocular camera is challenging, as it attempts to reason 3D information from 2D images. The connection between ground plane normals and ego-motion is initially modelled in HMM (Hidden Markov Model) [24], then the odometry and ground plane normals are jointly estimated from image sequences. Zhou et al. propose to use constrained homography to estimate the ground plane for robot platforms [23]. In terms of monocular Visual Odometry (VO), it is common to combine the ground plane estimation with scale recovery [40,41]. Our method is fundamentally different from the aforementioned methods: we decouple those multiple tasks and only estimate the normal vector from ego-motion with our specially designed IEKF. In such a way, our proposed

method supports monocular setup and other algorithms with different sensors, such as IMU-only odometry.

Recently, some Convolutional Neural Networks (CNN) have also been proposed to estimate ground planes. Particularly, given a monocular image sequence, photometric consistency can be used with homography warping to recover the normal vector in a self-supervised manner [25,26]. To further improve the accuracy, GroundNet [27] jointly learns pixel-level normals, ground segmentation, and depth maps using multiple networks. As a result, their latency is relatively high, ranging from 130 to 920 milliseconds/frame. While our method mainly focuses on the ground plane normal vector estimation using ego-motion, as detailed in Section 5, the latency is reduced to 3 milliseconds/frame for the IMU odometry and 50 milliseconds/frame for the monocular visual odometry. Moreover, the ground truth of ground plane normal is difficult to obtain and verify. Most existing works apply homography transformation on original images to produce augmented inputs and corresponding labels [24,27]. These methods consider normal vectors of original images as the fixed value calculated by extrinsic. However, in practice, such an approximation is inaccurate, and the augmentation deviates the data distribution from actual use cases. Instead, we use LiDAR points to calculate the ground plane normal as ground truth. The effectiveness is qualitatively and quantitatively verified in Section 5.

3. Ground Plane Normal

In this part, we structurally study the dynamics of the ground plane normal, thereby verifying the motivation behind this work. We argue that the ground plane normal vector in a vehicle's reference system is oscillating when the vehicle is moving. To verify it, we take a clip from KITTI [28] odometry sequence # 00 for illustration. Theoretically, if the ground plane normal remains constant, the IPM images (with fixed extrinsic) should be similar (e.g., the parallel road lanes and edges) between adjacent frames.

However, as visualized in Figure 3, the road edges between adjacent frames (with fixed extrinsic) are not well aligned after IPM with a constant ground plane normal. To explore this phenomenon, we use LiDAR points from the dataset to calculate the ground truth (GT) of the ground plane normal. Built on that, the GT road edges are marked in red dot lines. We clearly find that most real road edges are not properly aligned with GT, with more than 1° out of calibration. To get more general statistics of such dynamics, we count the number of frames based on their variation to the GT in roll and pitch. The final statistics are presented in Figure 4. It can be observed that the mean variation of pitch and roll angles are around 1.2° and 3.5° , respectively. In other words, rather than constant, the ground plane normal vector is dynamically changed when the vehicle is moving.

Similarly, Table 1 presents the mean values of pitch and roll dynamics on all KITTI odometry sequences. We can draw the same conclusion: the ground plane normal is not constant (around 1°) when a vehicle is moving. Such instability could further influence the performance of autonomous driving tasks. Therefore, our estimated ground plane normal vector (in Section 4) can be encoded to improve the robustness of autonomous driving applications.

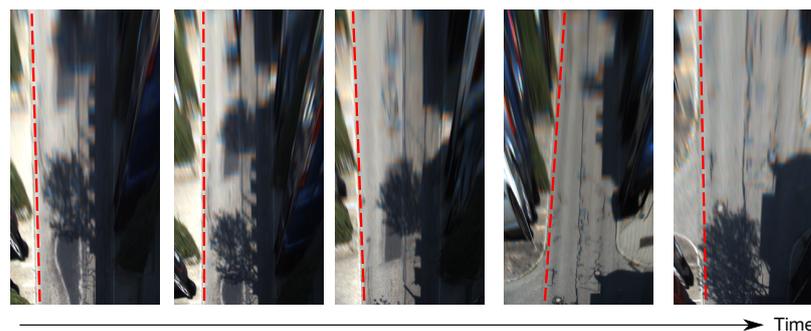


Figure 3. IPM images with the constant ground plane normal: road edges are not properly aligned.

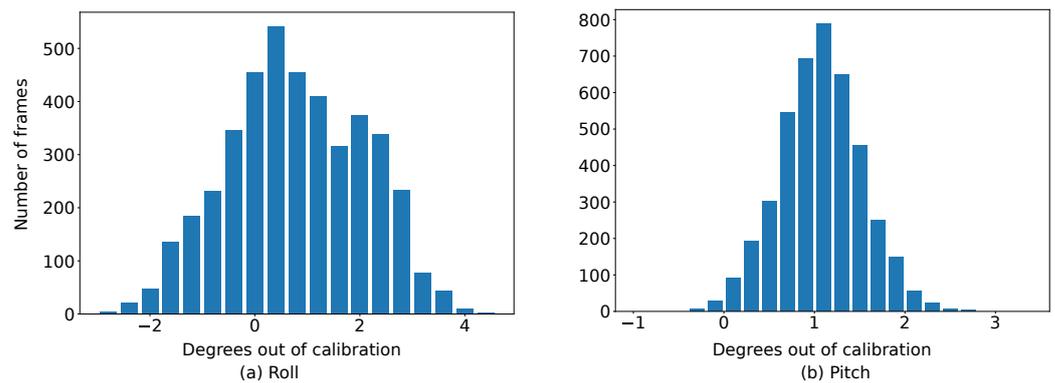


Figure 4. Statistics of frames (KITTI odometry sequence # 00) that are out of calibration in pitch and roll.

Table 1. Statistics of pitch and roll dynamics on all KITTI odometry sequences.

Sequence	00	01	02	03	04	05	06	07	08	09	Mean	Std
Pitch	1.06	1.16	1.11	0.40	1.21	1.27	1.27	1.27	1.31	1.47	1.15	0.27
Roll	0.92	0.59	1.20	1.30	1.46	0.99	0.78	0.70	0.93	0.91	0.98	0.26

4. Approach

In this part, our proposed ground plane normal estimation method is detailed. Figure 5 presents the pipeline. In short, we formulate the relationship between the odometry (from images or IMU) and ground plane normal based on IEKF. For a better description, We use a front-facing monocular camera on a wheeled vehicle as an example in the following descriptions.

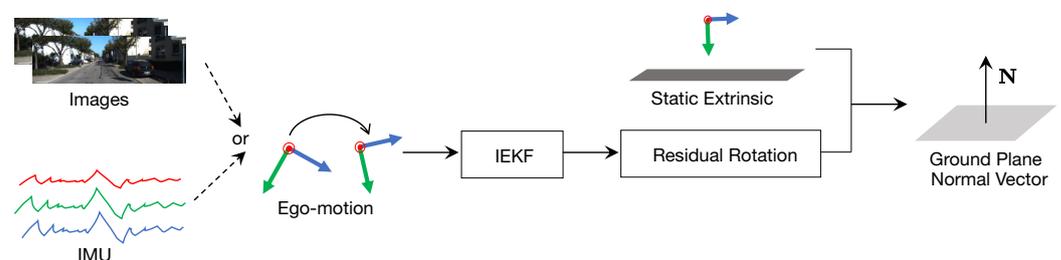


Figure 5. Overview of our proposed ground plane normal estimation pipeline. Our proposed IEKF can process ego-motion from various sensors, such as IMU, visual odometry from monocular images, and SLAM systems that can provide real-time odometry information. The final ground plane normal vector N is predicted in real-time based on the combination of residual rotation from IEKF and static extrinsic from prior calibration.

For a moving vehicle, its camera pose is trivially coherent with the ground plane. In real environments, the road surface is not ideally plane, but a segment close to the camera is approximately flat. In such a case, it is applicable to calculate the normal vector of the segment in the camera reference system. Specifically, when the vehicle is static, the ground plane normal vector can be computed from extrinsic parameters between the camera and the ground plane. The extrinsic can be easily obtained via off-line checkboard calibration [42]. When the vehicle is moving, due to oscillations of roll and pitch angles (see Figure 1), the extrinsic is no longer accurate to represent the relationship between the camera and the ground plane. In such a case, our proposed method is triggered. The rationale behind our method is that the dynamics of a normal vector can be roughly divided into two parts in the frequency domain: The low-frequency part describes the actual elevation changes, such as bumps and bridges. The high-frequency part is the oscillation, mainly because of braking and acceleration. Our goal is to split these two

components from ego-motion to calculate the ground plane normal vectors. In summary, our proposed method is built on two assumptions: (1) The close-to-camera road surface can be approximated as a flat plane. (2) The mean camera pose is close to its static extrinsic calibration.

Figure 6 presents the camera reference system of two adjacent frames. The transformation between the actual (vehicle moving, dark green and blue) and ideal (vehicle stopped, light green and blue) camera reference system is equal to the extrinsic rotation between the camera and ground plane. Accordingly, it can be used to calculate the ground plane normal vector:

$$N_k = \mathcal{N}(T_k^{-1} \cdot T'_k) \quad (1)$$

where $\mathcal{N}(\cdot)$ means extracting the second column (y-component) of the rotation from a transformation matrix. The rotation of $T_k \cdot T'_k$ can be decomposed to Euler angles: roll (z-axis), pitch (x-axis) and yaw (y-axis). For a moving vehicle, as shown in Figure 4, the pitch angle is the most dynamic component. Our task is to estimate pitch angle (θ_k in Figure 6), or more generally the residual rotation $T_k \cdot T'_k$:

$$T_{k-1}^k = T_k^{-1} \cdot T_{k-1}. \quad (2)$$

As T_k is available from ego-motion, the problem is now turned into estimating and tracking the ideal (vehicle stop) camera reference system T'_k . At first glance, this is trivial since T'_k is static and always parallel to the ground. However, the only input is ego-motion (the transformation between adjacent frames in the world reference system (WRS)). Even if the WRS is aligned with the ground plane, ego-motion unavoidably suffers from the drifting of long sequences. Thus, estimating T'_k from limit frames of history odometry is necessary, which intuitively leads to Kalman filter [43] as a potential solution.

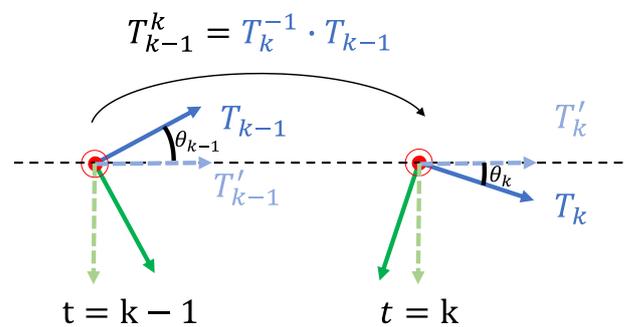


Figure 6. 2D side view of the camera reference system in two adjacent frames. T'_{k-1} and T'_k are the ideal camera reference system when the vehicle is stopped. T_{k-1} and T_k are the actual camera poses. $T_{k-1}^k = T_k^{-1} \cdot T_{k-1}$ is the ego-motion between two frames. The black dashed line is the ideal horizontal line parallel to the ground plane. θ_{k-1} and θ_k are the pitch angles relative to the ground plane. The actual camera extrinsics to the ground plane are $T_{k-1} \cdot T'_{k-1}$ and $T_k \cdot T'_k$, which is equivalent to the ground plane normal vector. Best view in colour.

To do so, we adopt the idea of IEKF [44,45] to our rotation estimation scenario. The general idea of IEKF is to use a deterministic non-linear observer directly on Lie groups instead of using a correction term on linear output. As shown in Figure 5, our method takes ego-motion as the input and output ground plane normal vector N . The source of ego-motion can be chosen from monocular SLAM system [46,47], learning-based monocular odometry [48–51], pure IMU-based odometry [52], and other SLAM (or odometry) algorithms that provide real-time ego-motion between frames.

The whole procedure of our proposed method is described in Algorithm 1. In terms of IEKF, it is adopted as follows: The state of the filter is a member of $SO(3)$, as we only consider the rotation of the sensor. The state and its covariance are initialized with zeros and identity matrix, respectively. We only consider a zero-order state ($SO(3)$), i.e., the process model is an identity function on the input rotation. Higher order state (e.g., angular velocity) could

be added to the filter if the source odometry sensor provides such observation, such as IMU. Nevertheless, we found that a constant process model is enough in our cases and makes our approach more general. If the sensor odometry provides relative transformation (i.e., T_k^{k-1}), the absolute transformation (i.e., T_k) is tracked by integration over time. The observation of the filter is the rotation part of T_k . To calculate the normal vector (N_i) of the current frame, the residual rotation (G_i) is calculated by the difference between the prediction state (Y_i) and absolute transformation (T_k). Note that the prediction state is calculated before the observation of the current frame is applied to the filter.

Algorithm 1 Ground Plane Normal Vector Estimation

Require: Extrinsic calibration between reference sensor and ground plane E_r^g

Input: Ego-motion from the reference sensor: $[T_0, T_1^0, \dots, T_k^{k-1}]$

Output: Ground plane normal vector w.r.t reference sensor: $[N_1, N_2, \dots, N_k]$

Initialization:

Covariance matrix $C = I_3$

Initial state $x \in SO(3)$

Process model $x = f(x)$

Process variance $P = p \cdot I_3$

Measurement model $\hat{x} = x$

Measurement variance $M = I_3$

Invariant Extended Kalman Filter $\mathcal{F}(x, C, P, M)$

Cumulative ego odometry $T_i \in SO(3)$

for $t = 0, 1, \dots, k$ **do**

 Compute $T_t = T_{t-1} \cdot T_t^{t-1}$

 Predict state: $T_t' = \mathcal{F}.predict()$

 Update filter: $\mathcal{F}.update(T_t)$

 Compute residual rotation: $G_t = T_t^{-1} \times T_t'$

 Compute normal vector N_t from residual rotation G_t using Equation (1)

end for

5. Experiments

In this part, we first introduce the implementation details of our proposed method. Built on that, we evaluate its performance quantitatively and qualitatively. Finally, the limitations of our method are discussed.

5.1. Implementation

To validate that the proposed method is agnostic to the source of ego-motion, we choose two challenging sensor setups for evaluation: monocular camera and pure IMU odometry. The experiments are conducted on the popular KITTI dataset [28]. It provides four front-facing camera images, raw IMU measurement data, LiDAR points, extrinsic calibration, and ground truth ego-motion. For monocular setup, ORB-SLAM2 [46] is applied on the left RGB camera images to obtain ego-motion. In terms of IMU-only odometry, the AI-IMU [52] is employed to extract ego-motion. After that, the extrinsic is used to convert the ego-motion from the IMU reference system to the camera reference. Note that KITTI provides IMU data at 100 Hz while the camera is running at 10 Hz. To fairly compare different odometry sources, the frame rate of IMU odometry is down-sampled to 10 Hz via integration.

To quantitatively evaluate our proposed method, the ground truth of the ground plane normal is calculated using LiDAR point clouds. Specifically, for each frame, the point cloud is projected onto the image to get 2D-3D correspondence, thereby selecting points within the camera's visual hull. Then, an off-the-shelf semantic segmentation method [53] is applied to images to obtain image masks for ground areas. Finally, the RANSAC [54]

plane fitting is applied on the points which only correspond to the image ground area. For IEKF, the scale of process variance p is set to 10^{-2} . All our experiments run on a desktop with an Intel i5-6600K CPU running at 3.50 GHz. The operating system is Ubuntu 18.04.6 LTS. Note that, unlike GroundNet [27], GPU is not required by our proposed method.

5.2. Quantitative Evaluation

Here, the estimated ground plane normal vectors are evaluated against ground truth:

$$E_{rad} = \frac{\sum_{i=1}^k \arccos(N_i^{est} \cdot N_i^{gt})}{k}, \quad (3)$$

where E_{rad} is vector errors in radians, N_i^{est} and N_i^{gt} are the estimated and ground truth vectors in i -th frame, respectively. All the normal vectors are unitary vectors with modulus = 1. As mentioned in Section 5.1, there are two types of ground truth: fixed extrinsic and plane fitting. For the first one, the ground truth normal vector is constant and calculated from calibration. For the second one, the ground truth normal vectors are calculated by the plane fitting from LiDAR points. For a fair comparison, we keep the original setting of existing methods and apply our method to both IMU and monocular sensors. Table 2 presents the detailed results. As presented in Table 2, our proposed method achieves the best accuracy in both sensors. For instance, the estimated vector error on the KITTI dataset is reduced from 3.02° with [26] to 0.39° with our proposed method. Moreover, the monocular-based method provides slightly better results compared with IMU-only odometry. This is because the accuracy of monocular odometry is inherently higher than IMU odometry. We also compare the computation time (if provided) in Table 2. It can be clearly observed that the computation time with our method is between 3–50 ms per frame, dramatically reduced as well. Overall, our proposed method can estimate accurate ground plane normal vectors in real time.

Table 2. Quantitative comparison of our proposed method with previous works. The running time is also compared here to demonstrate the improvement in efficiency using our method. Particularly, the adopted IEKF takes less than one millisecond per frame.

Methods	Error (°)	Time (ms/frame)
HMM [24]	4.10	-
Xiong [26]	3.02	-
GroundNet [27]	0.70	920
Road Aware [25]	1.12	130
Naive [54]	0.98	-
Ours (IMU)	0.44	3 = 2 (IMU odometry) + 1 (IEKF)
Ours (Monocular)	0.39	50 = 49 (Visual odometry) + 1 (IEKF)

5.3. Qualitative Evaluation

To better understand our contributions, the IPM images with static (from fixed extrinsic calibration) and dynamic (from our proposed method) normal vectors are visually compared in Figure 7. Here, static normal vector means the ground plane normals are constant [26]. Ideally, if the ground plane normals used in IPM are accurate, the parallel lanes on the flat road surface should maintain parallel in IPM images (see Section 1). However, as shown in Figure 7a, the road lanes are not properly parallel with the static normal vector. However, with dynamic normal vector from our method, the road edges in IPM images are more parallel and consistent in Figure 7b.

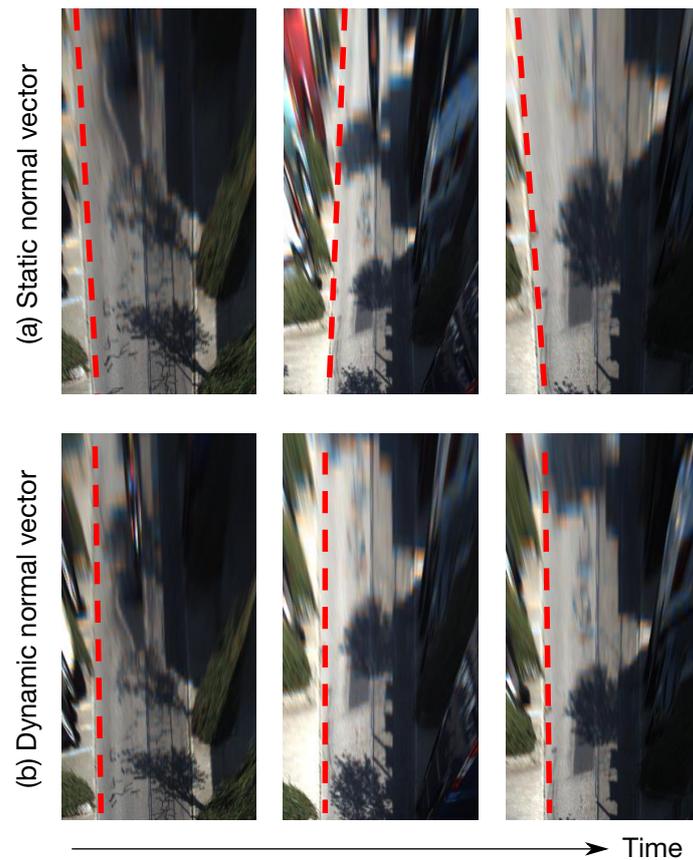


Figure 7. Visual comparison of IPM images using (a) static normal vector based on fixed extrinsic calibration, and (b) dynamic normal vector using our proposed method. The odometry input is formed by the monocular version of ORB-SLAM2. We can clearly find that the road edges are not parallel with each other with a static normal vector. Based on the dynamic normal vector from our method, the road edges in IPM images are more parallel and consistent.

Figure 8 details pitch angle variations in the IPM sequence of the KITTI odometry sequence-00 clip. Based on the dynamic normal vectors using our proposed method, we can clearly find that the pitch angles (both monocular camera and IMU) are properly aligned with ground truth among most frames. However, in some cases (frame 500 to 600), the estimated ground normals differ from the GT. The reason is that the vehicle is making a sharp right turn, and the proposed method with IEKF can not produce an ideal estimation under extreme vehicle dynamics. As discussed in [3–5], normal vector estimation is inherently equivalent to vanishing lines estimation. Thus, converting ground normals into vanishing lines (in original image space) can also provide convincing visualization of our proposed method. In Figure 9, the green line is calculated from our proposed method, showing a reasonable vanishing line estimation. The red line is calculated from static calibration (static normal vector) and apparently deviates from the ideal one. A better visualization can be found in the supplementary video.

To verify the robustness of our proposed method, we conduct the same experiments on the nuScenes [55] dataset. As shown in Figure 10, the images on the left are IPM results using origin fixed camera extrinsic. The images on the right show IPM results using ground plane normals estimated by our proposed methods. We can clearly see that the proposed method produces more stable and reasonable IPM images.

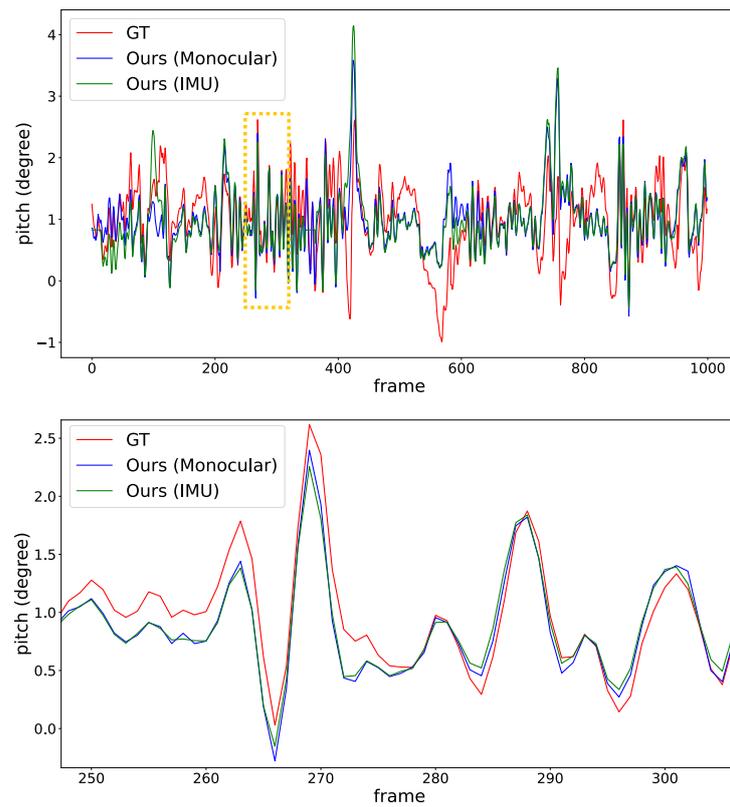


Figure 8. Plots of pitch angles with normal vectors calculated by the proposed methods. The bottom plot shows the details within 50 frames from the orange box. The oscillation tendency of the pitch angles from the proposed methods aligns well with the ground truth. Note that the overall amplitude of the pitch angles is actually small, usually within 1 degree.



Figure 9. Visualization of vanishing lines. The red and green horizontal lines are vanishing lines converted from fixed and dynamic ground plane normals, respectively. The **(bottom)** image is a zoom-in image of orange rectangular areas from the **(top)** image. The green line is obviously a more accurate estimation of the vanishing line.

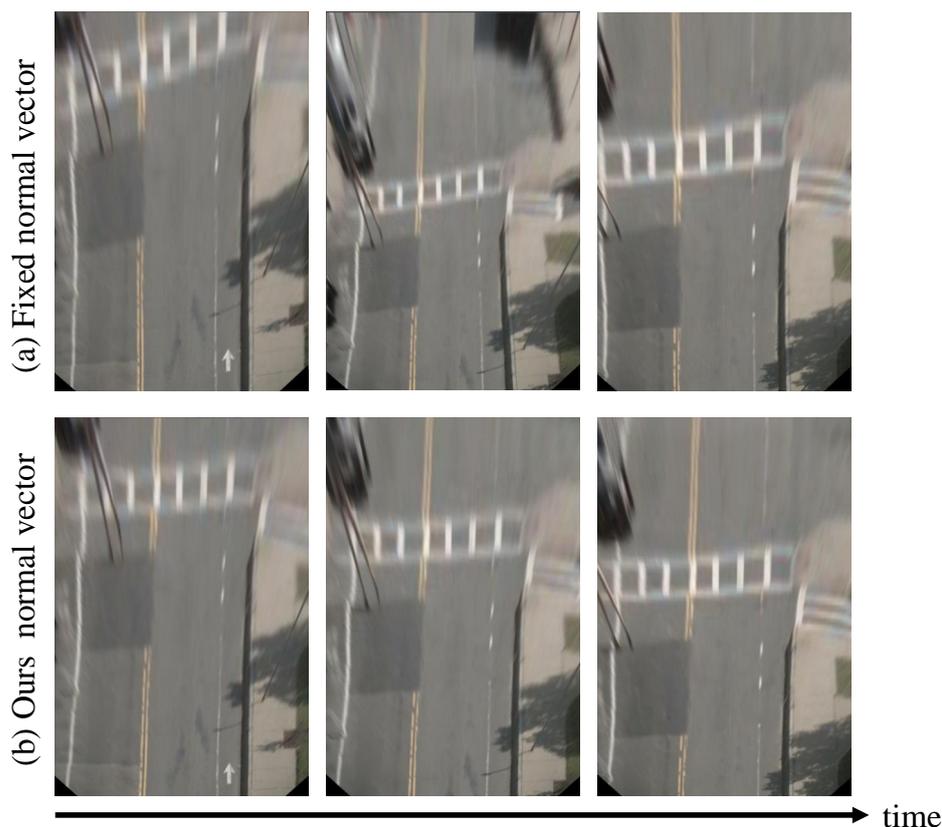


Figure 10. IPM visualization on the nuScenes dataset.

5.4. Ablation Study

To evaluate the effectiveness of using IEKF on the odometry to calculate the ground plane normal, we conduct extra experiments by using odometry only to obtain the ground plane normal. There are two ways to use odometry information directly: relative odometry and absolute odometry. The former is the relative pose between adjacent frames provided by the odometry algorithm, and the latter is accumulated odometry, i.e., current pose w.r.t first frame. As shown in the Figure 11, using pure relative odometry results in inconsistent ground normal estimation in some cases. This is because relative rotation between frames only contains “instant information” of the vehicle poses, thus being unable to handle various road surfaces such as small slopes or bumps. For absolute odometry, the result is even worse as it suffers from drifting issues as errors of odometry accumulate over time. Quantitative results are shown in Table 3.

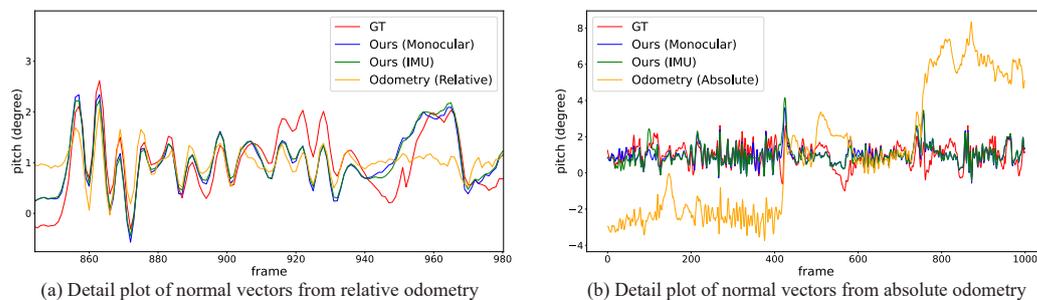


Figure 11. Comparing ground plane normal estimated by odometry only.

Table 3. Quantitative comparison ground plane normal estimation between our proposed methods and using odometry directly.

Methods	Error (°)
Pure odometry(relative)	1.09
Pure odometry(absolute)	2.98
Naive(constant normal)	0.98
Ours(IMU)	0.44
Ours(Monocular)	0.39

6. Limitations

Though our proposed method can estimate accurate ground plane normal vectors in real-time, there are still two limitations: (1) Our proposed method can only be applied in wheeled vehicles since it relies on the underlying connection between the ground plane and ego-motion in wheeled vehicles. (2) Our proposed method relies on the assumption that the nearby ground plane can always be approximated as a flat plane and the vehicle is driving smoothly. Thus, the estimation accuracy would degrade if the vehicle is driving on extremely uneven roads such as terrains and slopes or making harsh turns. For these cases, the effective range of the ground plane normal estimated by our proposed method could narrow down to smaller areas.

7. Conclusions

In this paper, we propose a ground plane normal vector estimation in driving scenes. We structurally study the dynamics of normal vectors when the vehicle is moving, which were previously considered constants. The argument is verified with both visualization and quantitative experiments. After analyzing the underlying connection between ground plane normals and vehicle odometry, the invariant extended Kalman filter is adopted to estimate the normal vectors with high accuracy in real time. The input of the filter is agnostic to the sensors that produce odometry information. Experiments on public datasets demonstrate that our method achieves promising accuracy on both monocular and IMU-only odometry.

Supplementary Materials: The following are available at <https://www.mdpi.com/article/10.3390/s22239375/s1>, supplementary video: IPM.

Author Contributions: Conceptualization, J.Z. and C.Y.; Data creation, J.Z.; Funding acquisition, Q.Z., T.C. and C.Y.; Methodology, J.Z.; Project administration, W.S.; Software, J.Z. and W.S.; Supervision, W.S. and C.Y.; Validation, W.S., Q.Z., T.C. and C.Y.; Writing—original draft, J.Z. and C.Y.; Writing—review & editing, C.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Key Research and Development Program of China (Grant Number: 2019YFB1310900), the National Natural Science Foundation of China (Grant Number: 62073229), and Jiangsu Policy Guidance Program (International Science and Technology Cooperation) The Belt and Road Initiative Innovative Cooperation Projects (Grant Number: BZ2021016), the Research Fund of Horizon Robotics, and The Natural Science Foundation of the Jiangsu Higher Education Institutions of China (Grant Number: 22KJB520008).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The research uses the KITTI (<https://www.cvlibs.net/datasets/kitti>) and the nuScenes (<https://www.nuscenes.org>) datasets (accessed on 28 November 2022).

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

References

1. Jazar, R.N. *Vehicle Dynamics*; Springer: Berlin, Germany, 2008; Volume 1.
2. Liu, T.; Liu, Y.; Tang, Z.; Hwang, J.N. Adaptive ground plane estimation for moving camera-based 3D object tracking. In Proceedings of the IEEE International Workshop on Multimedia Signal Processing, New Orleans, LA, USA, 24–26 November 2017; pp. 1–6.
3. Wang, Y.; Teoh, E.K.; Shen, D. Lane detection and tracking using B-Snake. *Image Vis. Comput.* **2004**, *22*, 269–280. [[CrossRef](#)]
4. Chen, Q.; Wang, H. A real-time lane detection algorithm based on a hyperbola-pair model. In Proceedings of the IEEE Intelligent Vehicles Symposium, Götterburg, Sweden, 19–22 June 2016; pp. 510–515.
5. Garnett, N.; Cohen, R.; Pe'er, T.; Lahav, R.; Levi, D. 3d-lanenet: End-to-end 3d multiple lane detection. In Proceedings of the IEEE International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 2921–2930.
6. Yang, C.; Indurkha, B.; See, J.; Grzegorzec, M. Towards automatic skeleton extraction with skeleton grafting. *IEEE Trans. Vis. Comput. Graph.* **2020**, *27*, 4520–4532. [[CrossRef](#)] [[PubMed](#)]
7. Qian, Y.; Dolan, J.M.; Yang, M. DLT-Net: Joint detection of drivable areas, lane lines, and traffic objects. *IEEE Trans. Intell. Transp. Syst.* **2019**, *21*, 4670–4679. [[CrossRef](#)]
8. Soquet, N.; Aubert, D.; Hautiere, N. Road segmentation supervised by an extended v-disparity algorithm for autonomous navigation. In Proceedings of the IEEE Intelligent Vehicles Symposium, Istanbul, Turkey, 13–15 June 2007; pp. 160–165.
9. Alvarez, J.M.; Gevers, T.; LeCun, Y.; Lopez, A.M. Road scene segmentation from a single image. In Proceedings of the European Conference on Computer Vision, Florence, Italy, 7–13 October 2012; pp. 376–389.
10. Lee, D.G. Fast Drivable Areas Estimation with Multi-Task Learning for Real-Time Autonomous Driving Assistant. *Appl. Sci.* **2021**, *11*, 10713. [[CrossRef](#)]
11. Lee, D.G.; Kim, Y.K. Joint Semantic Understanding with a Multilevel Branch for Driving Perception. *Appl. Sci.* **2022**, *12*, 2877. [[CrossRef](#)]
12. Knorr, M.; Niehsen, W.; Stiller, C. Online extrinsic multi-camera calibration using ground plane induced homographies. In Proceedings of the IEEE Intelligent Vehicles Symposium, Gold Coast City, Australia, 23–26 June 2013; pp. 236–241.
13. Yang, C.; Wang, W.; Zhang, Y.; Zhang, Z.; Shen, L.; Li, Y.; See, J. MLife: A lite framework for machine learning lifecycle initialization. *Mach. Learn.* **2021**, *110*, 2993–3013. [[CrossRef](#)] [[PubMed](#)]
14. Yang, C.; Yang, Z.; Li, W.; See, J. FatigueView: A Multi-Camera Video Dataset for Vision-Based Drowsiness Detection. *IEEE Trans. Intell. Transp. Syst.* **2022**. [[CrossRef](#)]
15. Liu, J.; Cao, L.; Li, Z.; Tang, X. Plane-based optimization for 3D object reconstruction from single line drawings. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *30*, 315–327.
16. Chen, X.; Kundu, K.; Zhang, Z.; Ma, H.; Fidler, S.; Urtasun, R. Monocular 3d object detection for autonomous driving. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2147–2156.
17. Qin, Z.; Li, X. MonoGround: Detecting Monocular 3D Objects From the Ground. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 19–20 June 2022; pp. 3793–3802.
18. Zhou, D.; Dai, Y.; Li, H. Ground-plane-based absolute scale estimation for monocular visual odometry. *IEEE Trans. Intell. Transp. Syst.* **2019**, *21*, 791–802. [[CrossRef](#)]
19. Qin, T.; Zheng, Y.; Chen, T.; Chen, Y.; Su, Q. A Light-Weight Semantic Map for Visual Localization towards Autonomous Driving. In Proceedings of the IEEE International Conference on Robotics and Automation, Xi'an, China, 30 May–5 June 2021; pp. 11248–11254.
20. Reiher, L.; Lampe, B.; Eckstein, L. A sim2real deep learning approach for the transformation of images from multiple vehicle-mounted cameras to a semantically segmented image in bird's eye view. In Proceedings of the IEEE International Conference on Intelligent Transportation Systems, Rhodes, Greece, 20–23 September 2020; pp. 1–7.
21. Phillion, J.; Fidler, S. Lift, splat, shoot: Encoding images from arbitrary camera rigs by implicitly unprojecting to 3d. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; pp. 194–210.
22. Li, Q.; Wang, Y.; Wang, Y.; Zhao, H. Hdmapnet: An online hd map construction and evaluation framework. In Proceedings of the IEEE International Conference on Robotics and Automation, Philadelphia, PA, USA, 23–27 May 2022; pp. 4628–4634.
23. Zhou, J.; Li, B. Robust ground plane detection with normalized homography in monocular sequences from a robot platform. In Proceedings of the International Conference on Image Processing, Atlanta, GA, USA, 8–11 October 2006; pp. 3017–3020.
24. Dragon, R.; Van Gool, L. Ground plane estimation using a hidden markov model. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 4026–4033.
25. Sui, W.; Chen, T.; Zhang, J.; Lu, J.; Zhang, Q. Road-aware Monocular Structure from Motion and Homography Estimation. *arXiv* **2021**, arXiv:2112.08635.
26. Xiong, L.; Wen, Y.; Huang, Y.; Zhao, J.; Tian, W. Joint Unsupervised Learning of Depth, Pose, Ground Normal Vector and Ground Segmentation by a Monocular Camera Sensor. *Sensors* **2020**, *20*, 3737. [[CrossRef](#)] [[PubMed](#)]
27. Man, Y.; Weng, X.; Li, X.; Kitani, K. GroundNet: Monocular ground plane normal estimation with geometric consistency. In Proceedings of the ACM International Conference on Multimedia, Nice, France, 21–25 October 2019; pp. 2170–2178.
28. Geiger, A.; Lenz, P.; Urtasun, R. Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012.

29. Gallo, O.; Manduchi, R.; Rafii, A. Robust curb and ramp detection for safe parking using the Canesta TOF camera. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Anchorage, Alaska, 23–28 June 2008; pp. 1–8.
30. Yu, H.; Zhu, J.; Wang, Y.; Jia, W.; Sun, M.; Tang, Y. Obstacle classification and 3D measurement in unstructured environments based on ToF cameras. *Sensors* **2014**, *14*, 10753–10782. [[CrossRef](#)] [[PubMed](#)]
31. Choi, S.; Park, J.; Byun, J.; Yu, W. Robust ground plane detection from 3D point clouds. In Proceedings of the International Conference on Control, Automation and Systems, Suwon si, Republic of Korea, 22–25 October 2014; pp. 1076–1081.
32. Zhang, W. Lidar-based road and road-edge detection. In Proceedings of the IEEE Intelligent Vehicles Symposium, La Jolla, CA, USA, 21–24 June 2010; pp. 845–848.
33. McDaniel, M.W.; Nishihata, T.; Brooks, C.A.; Iagnemma, K. Ground plane identification using LIDAR in forested environments. In Proceedings of the IEEE International Conference on Robotics and Automation, Anchorage, AK, USA, 3–7 May 2010; pp. 3831–3836.
34. Miadlicki, K.; Pajor, M.; Sakow, M. Ground plane estimation from sparse LIDAR data for loader crane sensor fusion system. In Proceedings of the International Conference on Methods and Models in Automation and Robotics, Międzyzdroje, Poland, 28–31 August 2017; pp. 717–722.
35. Lee, Y.H.; Leung, T.S.; Medioni, G. Real-time staircase detection from a wearable stereo system. In Proceedings of the International Conference on Pattern Recognition, Tsukuba, Japan, 11–15 November 2012; pp. 3770–3773.
36. Schwarze, T.; Lauer, M. Robust ground plane tracking in cluttered environments from egocentric stereo vision. In Proceedings of the IEEE International Conference on Robotics and Automation, Seattle, WA, USA, 26–30 May 2015; pp. 2442–2447.
37. Kusupati, U.; Cheng, S.; Chen, R.; Su, H. Normal assisted stereo depth estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 2189–2199.
38. Se, S.; Brady, M. Ground plane estimation, error analysis and applications. *Robot. Auton. Syst.* **2002**, *39*, 59–71. [[CrossRef](#)]
39. Chumerin, N.; Van Hulle, M. Ground plane estimation based on dense stereo disparity. In Proceedings of the International Conference on Neural Networks and Artificial Intelligence, Prague, Czech Republic, 3–6 September 2008; pp. 1–5.
40. Song, S.; Chandraker, M. Robust scale estimation in real-time monocular SFM for autonomous driving. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1566–1573.
41. Zhou, D.; Dai, Y.; Li, H. Reliable scale estimation and correction for monocular visual odometry. In Proceedings of the IEEE Intelligent Vehicles Symposium, Gotenburg, Sweden, 19–22 June 2016; pp. 490–495.
42. Hartley, R.; Zisserman, A. *Multiple View Geometry in Computer Vision*; Cambridge University Press, Cambridge, UK, 2003.
43. Kalman, R.E. A new approach to linear filtering and prediction problems. *J. Basic Eng. (Am. Soc. Mech. Eng.)* **1960**, *82*, 35–45. [[CrossRef](#)]
44. Bonnabel, S. Left-invariant extended Kalman filter and attitude estimation. In Proceedings of the IEEE Conference on Decision and Control, New Orleans, LA, USA, 12–14 December 2007; pp. 1027–1032.
45. Barrau, A.; Bonnabel, S. The invariant extended Kalman filter as a stable observer. *IEEE Trans. Autom. Control* **2016**, *62*, 1797–1812. [[CrossRef](#)]
46. Mur-Artal, R.; Montiel, J.M.M.; Tardos, J.D. ORB-SLAM: A versatile and accurate monocular SLAM system. *IEEE Trans. Robot.* **2015**, *31*, 1147–1163. [[CrossRef](#)]
47. Engel, J.; Koltun, V.; Cremers, D. Direct sparse odometry. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 611–625. [[CrossRef](#)] [[PubMed](#)]
48. Yang, N.; Stumberg, L.V.; Wang, R.; Cremers, D. D3vo: Deep depth, deep pose and deep uncertainty for monocular visual odometry. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 1281–1292.
49. Zhang, J.; Sui, W.; Wang, X.; Meng, W.; Zhu, H.; Zhang, Q. Deep online correction for monocular visual odometry. In Proceedings of the IEEE International Conference on Robotics and Automation, Xi’an, China, 30 May–5 June 2021; pp. 14396–14402.
50. Wagstaff, B.; Peretroukhin, V.; Kelly, J. On the Coupling of Depth and Egomotion Networks for Self-Supervised Structure from Motion. *IEEE Robot. Autom. Lett.* **2022**, *7*, 6766–6773. [[CrossRef](#)]
51. Zhang, S.; Zhang, J.; Tao, D. Towards Scale Consistent Monocular Visual Odometry by Learning from the Virtual World. In Proceedings of the IEEE International Conference on Robotics and Automation, Philadelphia, PA, USA, 23–27 May 2022; pp. 5601–5607.
52. Brossard, M.; Barrau, A.; Bonnabel, S. AI-IMU dead-reckoning. *IEEE Trans. Intell. Veh.* **2020**, *5*, 585–595. [[CrossRef](#)]
53. Cheng, B.; Misra, I.; Schwing, A.G.; Kirillov, A.; Girdhar, R. Masked-attention mask transformer for universal image segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 19–20 June 2022; pp. 1290–1299.
54. Fischler, M.A.; Bolles, R.C. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* **1981**, *24*, 381–395. [[CrossRef](#)]
55. Caesar, H.; Bankiti, V.; Lang, A.H.; Vora, S.; Liong, V.E.; Xu, Q.; Krishnan, A.; Pan, Y.; Baldan, G.; Beijbom, O. nuScenes: A multimodal dataset for autonomous driving. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020.