remote sensing

MDPI

Article

Autonomous Vehicle Localization with Prior Visual Point Cloud Map Constraints in GNSS-Challenged Environments

Xiaohu Lin [1], Fuhong Wang [1,*], Bisheng Yang [2] and Wanwei Zhang [1]

[1] School of Geodesy and Geomatics, Wuhan University, Wuhan 430079, China; xhlin214@whu.edu.cn (X.L.); wwzhang@sgg.whu.edu.cn (W.Z.)
[2] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China; bshyang@whu.edu.cn
* Correspondence: fhwang@sgg.whu.edu.cn; Tel.: +86-027-6875-8529

Abstract: Accurate vehicle ego-localization is key for autonomous vehicles to complete high-level navigation tasks. The state-of-the-art localization methods adopt visual and light detection and ranging (LiDAR) simultaneous localization and mapping (SLAM) to estimate the position of the vehicle. However, both of them may suffer from error accumulation due to long-term running without loop optimization or prior constraints. Actually, the vehicle cannot always return to the revisited location, which will cause errors to accumulate in Global Navigation Satellite System (GNSS)-challenged environments. To solve this problem, we proposed a novel localization method with prior dense visual point cloud map constraints generated by a stereo camera. Firstly, the semi-global-block-matching (SGBM) algorithm is adopted to estimate the visual point cloud of each frame and stereo visual odometry is used to provide the initial position for the current visual point cloud. Secondly, multiple filtering and adaptive prior map segmentation are performed on the prior dense visual point cloud map for fast matching and localization. Then, the current visual point cloud is matched with the candidate sub-map by normal distribution transformation (NDT). Finally, the matching result is used to update pose prediction based on the last frame for accurate localization. Comprehensive experiments were undertaken to validate the proposed method, showing that the root mean square errors (RMSEs) of translation and rotation are less than 5.59 m and 0.08°, respectively.

Keywords: autonomous vehicle; stereo visual odometry; pose prediction; accurate vehicle localization

Citation: Lin, X.; Wang, F.; Yang, B.; Zhang, W. Autonomous Vehicle Localization with Prior Visual Point Cloud Map Constraints in GNSS-Challenged Environments. *Remote Sens.* **2021**, *13*, 506. https://doi.org/10.3390/rs13030506

Academic Editors: Nereida Rodriguez-Alvarez and Kai-Wei Chiang
Received: 27 November 2020
Accepted: 28 January 2021
Published: 31 January 2021

1. Introduction

Autonomous vehicles, location based services (LBSs), and mobile mapping systems (MMSs) in unknown environments are important, well researched, and active fields. In particular, the autonomous vehicle is considered the next revolutionary technology that will change people's lives in many ways [1].

Traditionally, autonomous vehicles rely on Global Navigation Satellite System, inertial navigation system (GNSS/INS) for localization [2,3], in which GNSS provides a drift-free position and then combines with the high-frequency relative pose estimated from INS for accurate localization. However, the differential signal cannot be received everywhere and GNSS position information is prone to jump because of signal blockage, multi-path effects, and magnetic noise [4,5], and INS may suffer from cumulative accumulated errors. Especially in GNSS-challenged environments, such as urban canyons, boulevards, and indoor environments, where accurate localization is limited. Considering the factors mentioned above, it is necessary to design a more accurate, stable and economical localization system for autonomous vehicles.

State-of-the-art localization methods adopt visual and LiDAR simultaneous localization and mapping (SLAM) to estimate the position of the vehicle. Visual SLAM can be classified as the feature-based method and the direct method, according to the adopted error model. The feature-based method is accurate and robust in a static environment

with rich texture, but can easily fail in dynamic environments with weak texture and pure rotation. The direct method makes up for some defects of the feature-based method, but most of them are limited to indoor environments. The LiDAR SLAM can make up for visual SLAM defects in weak texture and illumination changes. However, LiDAR is more expensive and less common than a camera. Moreover, SLAM may suffer from error accumulation due to long time running without a loop or prior constraints. Furthermore, the vehicle cannot always return to the revisited location, which is unable to perform loop-constrained optimization in GNSS-challenged environments. In recent years, the fast development of computer vision and MMS makes high-definition (HD) maps available [6,7]. Therefore, localization based on a prior map is promising. It requires registration between the map and onboard sensors. The most obviously favored method is to use the same sensor for mapping and localization. This line of research can be found in [8–10], in which the LiDAR was used for three-dimensional (3D) mapping and localization. However, researchers prefer to achieve localization performance using a camera due to the associated costs and physical requirements [11–15].

Currently, most map-based localization schemes involve mapping with LiDAR, and then the image collected by the camera is used for localization. However, the field of view (FOV) and resolution will bring challenges to the camera localization within a prior LiDAR map. In this paper—which is different from camera localization methods based on prior LiDAR maps—we developed a novel vehicle localization pipeline with a stereo camera and a priori visual point cloud map for autonomous vehicle localization, verified through the Karlsruhe Institute of Technology and Toyota Technological Institute (KITTI) datasets [16], and field test data with qualitative and quantitative analysis.

The main contributions of this paper are as follows: (1) we proposed a novel vehicle localization pipeline with a stereo camera and a priori visual point cloud map, which adopted the semi-global-block-matching (SGBM) algorithm to estimate the visual point cloud of each image frame. Stereo visual odometry is used to provide the initial position for the current visual point cloud. (2) Multiple filtering and adaptive priori map segmentation are conducted for fast matching and localization. (3) The current visual point cloud frame is matched with the candidate sub-map by normal distribution transformation (NDT), to update pose prediction based on the last frame for accurate localization, which overcomes error accumulation due to long-running without loop optimization. Moreover, comprehensive experiments verified the proposed method.

The remainder of this paper is organized as follows. Section 2 provides an overview of the related work. The proposed method is then elaborated in Section 3. Section 4 presents the experiments undertaken to evaluate the localization performance of the proposed method, after which a discussion is presented in Section 5. Finally, in Section 6, the conclusions and future work are presented.

2. Related Work

Map-based vehicle localization has been investigated, and extensive algorithms have been proposed, including LiDAR-based methods, LiDAR-camera-based methods, and camera-based methods.

LiDAR-based methods: it can be divided into LiDAR SLAM (LSLAM) and map-based localization techniques. LSLAM may suffer from error accumulation due to long time running without a loop [17]. To this end, the location should be updated frequently by global features. For example, building corners are used to rectify the error accumulation in [18]. The map-based localization techniques are promoted by easily available HD maps. In most cases, the point cloud is used as the prior map, and the vehicle localization is carried out by current LiDAR scan matching with the prior map. Levinson et al. [19] integrated Global Positioning System (GPS), Inertial Measurement Unit (IMU), wheel odometry, LiDAR data to generate the map offline, and proposed a technique for an accurate localization of moving vehicle with the particle filter (PF), based on maps of urban environments. Later, they extended their work to a probabilistic approach [20].

Kim et al. [21] proposed a method of lane map building and localization for autonomous vehicles using 2D laser rangefinder, in which lane markers were used as local features extracted from reflectivity values of the LiDAR scans. Qin et al. [12] introduced a Monte Carlo Localization (MCL)-based method utilizing the curb-intersection features on urban roads, and the road observation was fused with odometry information. However, LiDAR is more expensive and less common than cameras.

LiDAR-camera-based methods: recent camera localization in 3D LiDAR maps have paid more attention to problem solving caused by the inherent differences in cameras and LiDAR. Photometric matching is one of the most common approaches. Wolcott and Eustice [11] proposed a monocular camera localization method within a prior 3D map generated by a survey vehicle equipped with 3D LiDAR, which used LiDAR reflectance images and matched them to image intensity. The synthesized image with the maximum Normalized Mutual Information (NMI) was selected as the corresponding camera pose. Instead of synthesizing candidate LiDAR images, Stewart and Newman [22] introduced Normalized Information Distance (NID) to match the image with LiDAR intensities for camera pose estimation. Another strategy involving camera localization in the 3D map is based on geometry. Forster et al. [23] proposed a new method of registering 3D maps reconstructed from a depth sensor on the ground robot and a monocular camera from a micro aerial vehicle (MAV). Camera localization was achieved from scan matching between the 3D map and the point cloud reconstructed by bundle adjustment. Similarly, Caselitz et al. [24] used a visual odometry system to reconstruct sparse 3D points from image features, these points are continuously matched against the map to estimate the camera pose in an online fashion. The major problem of camera localization in prior LiDAR maps originates from different ways of data collection in the camera and LiDAR. The depth from stereo matching is consistent with LiDAR, but there are still differences, especially at the edges. Moreover, data density obtained from the two sensors are different. For example, LiDAR usually has a higher ranging accuracy, a wider horizontal FOV, and is vertically sparse, while a camera provides evenly distributed density.

Camera-based methods: researchers prefer to achieve the autonomous vehicle localization with a camera due to the associated costs and physical requirements. Brubaker et al. [25] proposed an affordable solution to vehicle self-localization using visual odometry and road maps. They introduced a probabilistic model and visual odometry measurements to estimate the car displacement relative to the road maps. However, prior information (e.g., speed limits and street names) need to be expected. Ziegler et al. [26] proposed a localization method used by an autonomous vehicle on the historic Bertha-Benz Memorial Route, which introduced detailed geometric maps to supplement its online perception systems, and two complementary visual localization techniques were developed based on point features and lane features. However, the map creation is partially a manual process and some specific methods need to patch-up incomplete mapping data. Radwan et al. [27] proposed a localization method based on textual features in urban environments, which used off-the-shelf text extraction techniques to identify text labels, and an MCL algorithm was introduced to integrate multiple observations. However, the text-spotting and data association approaches need further research. Spangenberg et al. [28] proposed a new localization method based on pole-like landmarks, because they are distinctive, stable for a long-term, and can be detected reliably by a stereo camera. Then, localization is performed by a PF approach coupled with a Kalman filter (KF) for robust sensor fusion. However, this method is only applicable to specific scenarios. Lyrio et al. [29] and Oliveira et al. [30] employed the neural network to perform autonomous vehicle localization, which includes correlating camera images and associated global positions. The neural network was used to build a representation of the environment in the mapping phase. Then, current images were used to estimate global positions based on previously acquired knowledge provided by the neural network map. However, the drawbacks of their method included unreliable initialization and being time-consuming. In this paper, we proposed a novel localization

method for an autonomous vehicle with prior visual point cloud constraints, which could reduce localization costs and improve accuracy and stability.

3. Methodology

The proposed pipeline of autonomous vehicle localization is illustrated in Figure 1. In the preprocessing, the internal parameters of a stereo camera were calibrated and stereo images were rectified to facilitate the stereo matching. Prior visual point cloud maps were obtained by dense 3D surface reconstruction of a large-scale streetscape from vehicle-borne imagery and LiDAR; note that LiDAR was only used for pose estimation in prior visual point cloud map generation, which needed multiple filtering and adaptive priori map segmentation for fast matching and localization. After which, the classical SGBM algorithm was adopted to estimate the current visual point cloud of each frame, and stereo visual odometry was used to provide the initial estimation of the current visual point cloud. Then, the current visual point cloud frame was matched with the candidate sub-map by NDT. Finally, the matching result was used to update the pose prediction based on the last frame for accurate localization.

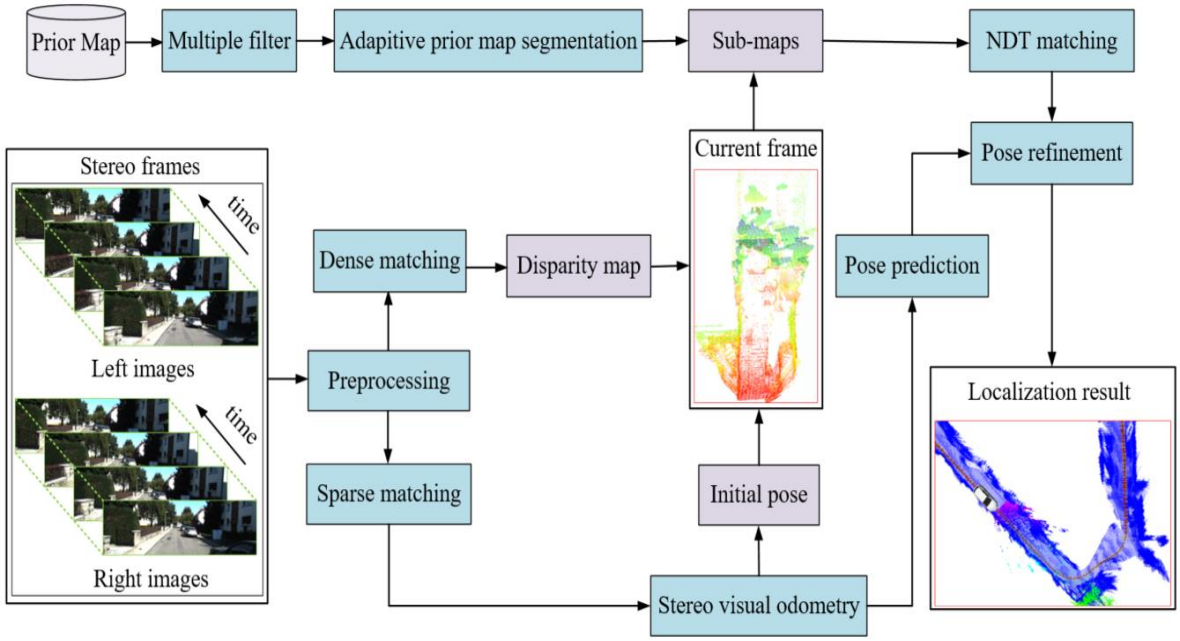

Figure 1. The proposed pipeline of autonomous vehicle localization.

3.1. Visual Point Cloud Generation

3.1.1. Priori Visual Point Cloud Map Generation

Priori visual point cloud maps were generated by dense 3D surface reconstructions of large-scale streetscapes from vehicle-borne imagery and LiDAR [31], which combined visual and laser odometry for robust and accurate pose estimation. The state-of-the-art pyramid stereo matching network (PSMNet) was used for estimating the depth information of stereo images [32]. Then, coarse-to-fine incremental dense 3D reconstruction was carried out by a key-frame selection [33] and coarse registration with local optimization using the iterative closest point (ICP) between the visual point cloud frames. Finally, a large number of redundant and noise points were removed through multiple filtering to further polish the reconstruction results.

3.1.2. Multiple Filtering

To reduce data redundancy and improve the efficiency of autonomous vehicle localization, the voxel grid filtering is introduced to downsample the visual point cloud, which

used the centroid in each tiny voxel grid to replace all points; to prevent the destruction of geometric structure information, the voxel grid size was set to $5 \times 5 \times 5$ cm^3. Meanwhile, the statistical outlier filtering (SOF) was used to remove noise points, which filtered outliers by computing the distance distribution of the point to its K-nearest neighbor (KNN) points. All points whose mean distances outside the range defined by global distance mean and standard deviation were marked as outliers and removed from the dataset by assuming that the results conformed to the Gaussian distribution [34]. In this paper, the number of KNN points to analyze for each visual point cloud was set to 200, and the standard deviation multiplier was set to 2, which means that all points whose distances were more than twice the standard deviation of the mean distance to the query point were considered as outliers and removed.

3.1.3. Adaptive Prior Map Segmentation

It is time-consuming to match the visual point cloud generated by a stereo camera with the whole prior visual point cloud map. The same is true for every time the vehicle moves, re-segmenting the prior map. An efficient management strategy of a grid map was proposed for real-time operation [35]. Similarly, adaptive prior map segmentation was performed on the prior map in this paper. The main idea was to segment the sub-map from a large point cloud, according to the size of a cube. The sub-map can be appropriately larger, and it will be re-segmented when the vehicle is about to leave the sub-map. The principle of segmentation is as follows:

$$\begin{cases} If, \left| L_{vechicle} - B_{edge} \right| > \tau, continue \\ else, B_{center} = L_{vechicle} \end{cases} \quad (1)$$

where $L_{vechicle}$ is the position of the vehicle, B_{edge} is edge of the sub-map, τ is the segmentation threshold, we set it to be 80 m in this paper. B_{center} is the new center point coordinates of the sub-map. The diagram of adaptive priori map segmentation from top view is as follows in Figure 2.

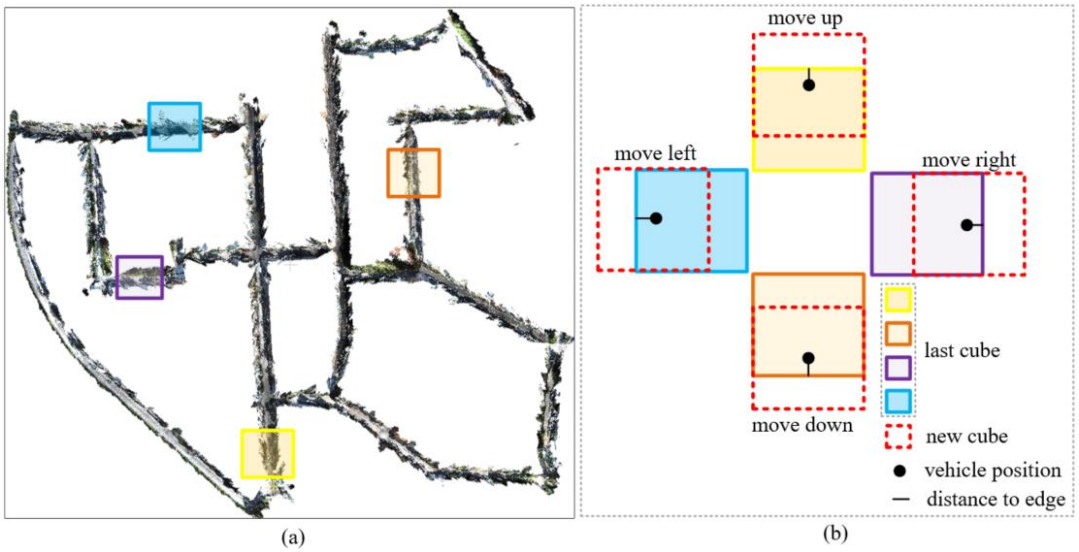

Figure 2. The diagram of adaptive priori map segmentation from top view. (**a**) Adaptive priori map segmentation from sequence 00 of the KITTI dataset; (**b**) different motion conditions for the vehicle.

Figure 2a shows adaptive priori map segmentation from sequence 00 of the KITTI dataset, and Figure 2b shows different motion conditions for the vehicle. After segmentation, the current visual point cloud is matched with the candidate sub-map by NDT to update pose prediction based on the last frame for accurate localization.

3.2. Current Visual Point Cloud Generation

Accurate and efficient matching from a pair of current stereo stream is of vital importance for the upcoming localization. The classic SGBM algorithm is adopted to estimate the disparity map, which is a stereo matching technique that estimates the optimal disparity of each pixel by minimizing an energy function. The whole process includes four steps: preprocessing, cost calculation, dynamic planning, and post-processing. Moreover, the energy function $E(d)$ is composed of mutual information based pixel-wise cost and global smoothness cost [36], as follows:

$$E(d) = \sum_{p}(C(p, d_p) + \sum_{q \in N_p} P_1 T\left[\left|d_p - d_q\right| = 1\right] + \sum_{q \in N_p} P_2 T\left[\left|d_{\mathrm{p}} - d_{\mathrm{q}}\right| > 1\right]) \tag{2}$$

where p and q are specific pixel values in the image, and d is the disparity map, N_p refers to the 8-neighborhood pixels of pixel p. Moreover, the function $T[\cdot]$ is defined to return 1 if its argument is true and 0 otherwise. The first term is the sum of all pixel matching cost for the disparities of d. The second term adds a constant penalty P_1 for all pixels q in the neighborhood N_p of p, for which the disparity changes a little bit. The third term adds a larger constant penalty P_2, for all larger disparity changes. Using a lower penalty for small changes permits an adaptation to slanted or curved surfaces. The constant penalty for all larger changes preserves discontinuities, which are visible as intensity changes. This is exploited by adapting P_2 to the intensity gradient. With the disparity map, the visual point cloud is generated as follows [37]:

$$Z\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & c_x \\ 0 & f & c_y \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} \tag{3}$$

$$\begin{cases} Z = \frac{bf}{d} \\ X = Z\frac{u - c_x}{f} \\ Y = Z\frac{v - c_y}{f} \end{cases} \tag{4}$$

where (u, v) is the image coordinate,f is the focal length, (c_x, c_y) is the principal point, b is the baseline, d is the disparity, and $(\mathrm{X}, \mathrm{Y}, \mathrm{Z})$ is the visual point cloud coordinate. The generated visual point cloud frames by dense stereo matching are shown in Figure 3, where a, b, c are from sequences 00, 05, 09 of the KITTI dataset, and d, e are from field test data Site-1 and Site-2.

3.3. Stereo Camera Localization with Prior Map Constraints

3.3.1. Initialization with Stereo Visual Odometry

Reliable initial pose estimation is the first step for autonomous vehicle localization with prior visual point cloud constraints. Stereo visual odometry provides an initial guess for the matching and localization of current visual point cloud with the prior visual point cloud. After calibration and rectification of stereo sequences $S = \left\{f_{k,l}, f_{k,r}\right\}_{k=1}^{n}$, where n represents the number of frames, the relative pose $T_{i,j} = \left\{R_{i,j}, t_{i,j}\right\} \in \mathbb{R}^{3\times4}$ between frames are estimated by stereo visual odometry based on the state-of-the-art Oriented FAST and Rotated BRIEF (ORB)-SLAM2 framework [38], where i represents the last frame and j represents the current frame, $R_{i,j} \in \mathbb{R}^{3\times3}$ is rotation matrix and $t_{i,j} \in \mathbb{R}^3$ is translation vector. To get accurate pose estimation between sequence image frames, local bundle adjustment (BA) on the key frames of visual odometry in the local window and global BA on all key frames are performed based on the Levenberg–Marquardt (L–M) method implemented in G2O [39].

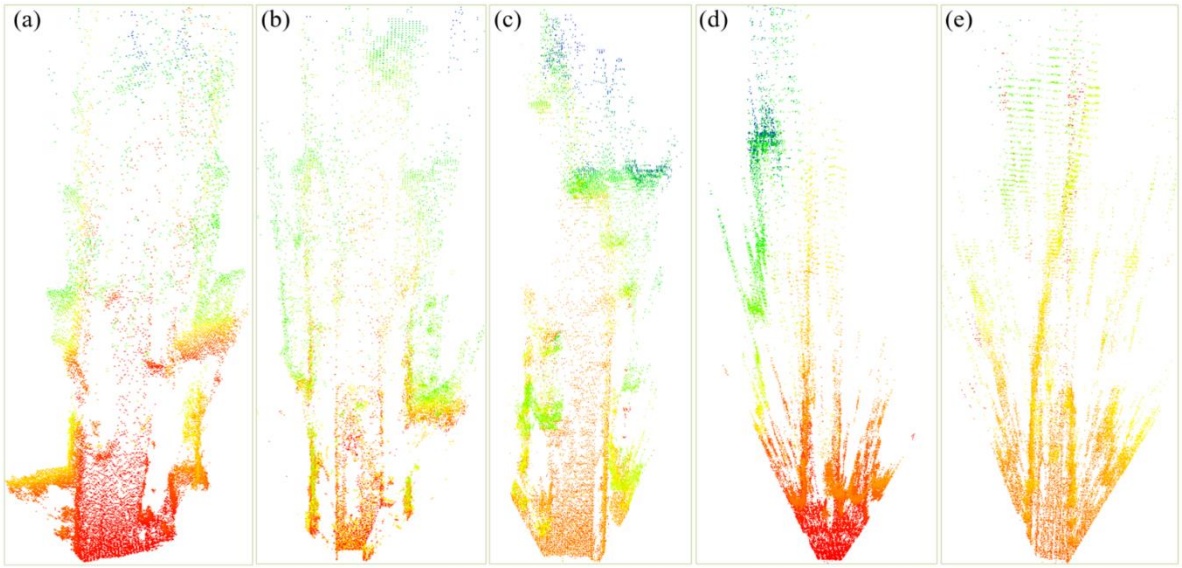

Figure 3. Visual point cloud frames generated by stereo matching. (**a**) The visual point cloud frame from sequence 00 of the KITTI dataset; (**b**) the visual point cloud frame from sequence 05 of KITTI dataset; (**c**) the visual point cloud frame from sequence 09 of KITTI dataset; (**d**) the visual point cloud frame from field test data Site-1; (**e**) the visual point cloud frame from field test data Site-2.

3.3.2. NDT Matching and Localization Refinement

With reliable initial pose estimation, current visual point cloud frames obtained from a stereo camera is matched with the candidate sub-map from the prior visual point cloud map to update pose prediction based on the last frame for accurate localization. Due to the good performance of NDT in precision, efficiency, and robustness to subtle changes [40,41], we selected NDT for matching. The main process of NDT matching here is to find the transformation parameters that maximize the likelihood of the current visual point cloud that lie on the candidate sub-map, or minimize the negative log-likelihood. If the transformation parameters can make the two scans match well, the probability density of the transformed points in the candidate sub-map will be large. The negative log-likelihood of the normal distribution grows without bounds for the point far from the mean. Thus, a mixture of normal and uniform distributions is considered [42]:

$$\overline{p}(\vec{x}) = c_1 \exp\left(-\frac{(\vec{x}-\vec{\mu})^T \Sigma^{-1} (\vec{x}-\vec{\mu})}{2}\right) + c_2 p_o \tag{5}$$

where $\vec{\mu}$ is the mean and Σ is the covariance, p_o is the expected ratio of outliers, we set it to 0.3. The calculation of c_1 and c_2 is shown in Equation (6), where $\Re$ is resolution and the default setting is 1. They are stable in our experiment. The function $\overline{p}(x) = -\log(c_1 \exp(-x^2/(2\sigma^2)) + c_2)$ can be approximated by a Gaussian $\widetilde{p}(x) = d_1 \exp(-d_2 x^2/(2\sigma^2)) + d_3$, Parameters d_1, d_2, d_3 are fitted by requiring that $\widetilde{p}(x)$ should behave like $\overline{p}(x)$ for $x = 0$, $x = \sigma$, and $x = \infty$:

$$\begin{cases} c_1 = 10(1 - p_o) \\ c_2 = p_o / \Re^3 \\ d_3 = -\log(c_2) \\ d_1 = -\log(c_1 + c_2) - d_3 \\ d_2 = -2\log((-\log(c_1 \exp(-0.5) + c_2) - d_3)/d_1) \end{cases} \tag{6}$$

Since all points are represented as Gaussian distributions, NDT is insensitive to uneven sampling distributions. We adopt the point-to-distribution (P2D) variant of NDT, which formulates the matching of a source cloud P_s to the target point cloud P_t, as a problem

of fitting the source points to the target's distribution. Then the best transformation parameters R_s^t, t_s^t is obtained by optimizing the NDT score function, as follows:

$$F(P_s, P_t, R_s^t, t_s^t) = \sum_{i=1}^{|P_s|} -d_1 \exp\left(-\frac{d_2(R_s^t p_{si} + t_s^t - \mu_i)^T \Sigma_i^{-1}(R_s^t p_{si} + t_s^t - \mu_i)}{2}\right) \quad (7)$$

where p_{si} is a source point in P_s, μ_i, Σ_i are the mean and covariance matrix of the NDT in which p_{si} lies. Special care is taken to the inverse of the covariance matrix Σ^{-1}. In case the points in a cell are perfectly coplanar or collinear, the covariance matrix is singular and cannot be inverted. Then Newton's iterative algorithm is used to find the best transformation parameters R_s^t, t_s^t.

Starting with a good initial value $P_{initial}$ provided by stereo visual odometry, the localization refinement is conducted by relative transformation parameters $T_{NDT}(k)$, which is obtained by NDT from current visual point cloud frame matched with the candidate sub-map. Suppose the last state is $T(k-1)$, the current pose $T(k)$ is updated by $T_{NDT}(k)$. The procedure of autonomous vehicle localization is described in Algorithm 1.

Algorithm 1. The procedure of localization with prior visual point cloud map constraints

Input: prior dense visual point cloud map M_{prior}, sequence stereo images $S = \{f_{k,l}, f_{k,r}\}_{k=1}^{n}$ at 10 Hz.
Output: refined pose $T = \{T(k)|k = 1, 2, \cdots, \mathbb{N}\}$ between sequence frames, where $T(k) = \{R_k, t_k\}$.
1: data preprocessing to get compact priori visual point cloud map M'_{prior};
2: multiple filtering and adaptive priori maps segmentation are performed on M'_{prior};
3: **for** each pair of stereo images S_k
4: **do**
5: rectified each pair of stereo images S_k to get S'_k;
6: dense stereo matching by SGBM to generate disparity map $M_{disparity}$;
7: convert disparity map $M_{disparity}$ to visual point cloud M_{visual} by formula (3-4);
8: estimate initial pose $P_{initial}$ from stereo visual odometry based on ORB-SLAM2 framework;
9: **If** $P_{initial}$ is initialized
10: make pose prediction based on the last frame $T(k-1)$;
11: perform NDT matching between the current visual point cloud and the candidate sub-map to get the relative transformation $T_{NDT}(k)$;
12: **else**
13: reinitialize $P_{initial}$;
14: **end for**
15: update current pose $T(k)$ by relative transformation parameters $T_{NDT}(k)$;
16: **return**
17: refined pose between frames $T = \{T(k)|k = 1, 2, \cdots, \mathbb{N}\}$.

4. Experimental Results

4.1. Experimental Platform Configuration

To evaluate the performance of our localization method, the KITTI dataset and field test experiments were carried out respectively. The laser scanner used in the KITTI dataset is Velodyne HDL-64E and the camera is Point Grey Flea 2. Our recording platform is the ground mobile vehicle, equipped with Velodyne VLP-16 and Point Grey GS3-PGE-23S6C-C camera, the experimental resolution of the camera is 1280 × 640 and the baseline is 0.495 m, as shown in Figure 4a. Both the laser scanner and stereo camera have acquisition frequencies of 10 Hz. The data collection environment includes both urban areas and suburbs, in which traditional methods are challenging, as shown in Figure 4b. The KITTI dataset is one of the most famous computer vision datasets, which provides with the sequences stereo images and high-precision reference trajectory. The field test data are two sequences of vehicle-borne image recorded by stereo camera. We employed state-of-the-art ORB-SLAM2 to generate the ground truth that served as a reference trajectory for the field test

experimental evaluation. Moreover, we made qualitative and quantitative evaluations from comparison with the reference trajectory, root mean square error (RMSE), and other methods. The computer used in our experiment is Intel Core i7-9700 @ 3.00 GHz and 32 GB RAM.

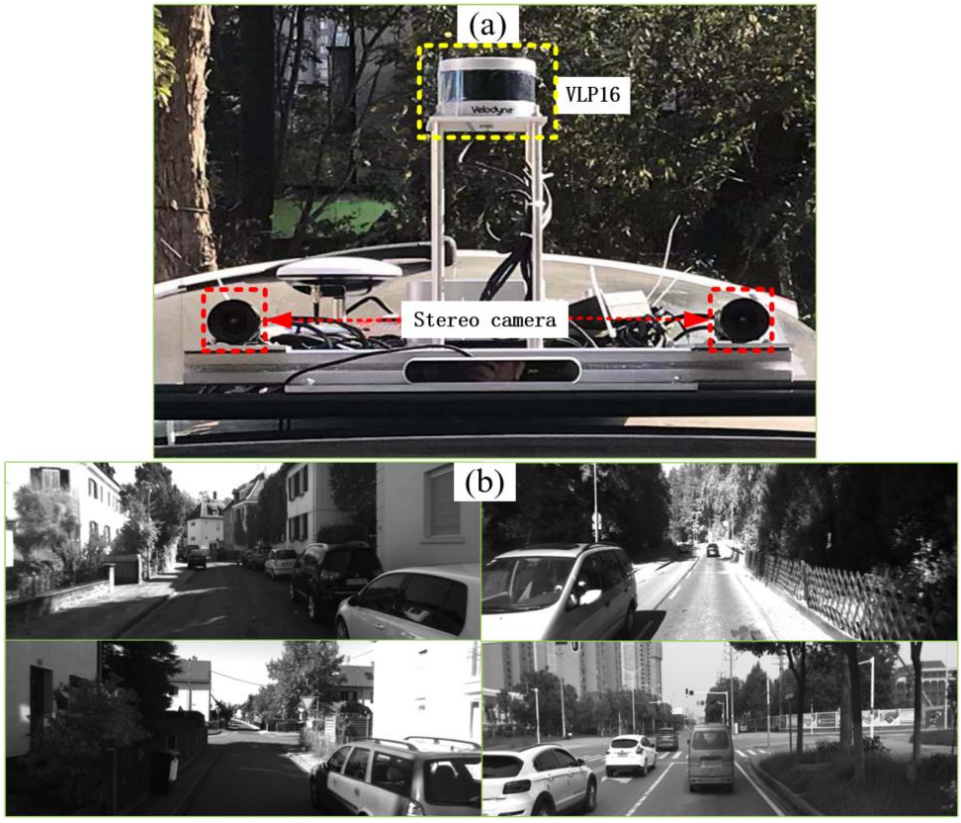

Figure 4. Data recording platform and corresponding collection environment. (**a**) Data recording platform; (**b**) data collection environment.

4.2. Qualitative Analysis

4.2.1. Evaluation with KITTI Dataset

The sequences 00, 05, 09 of the KITTI dataset were selected for experimental analysis. Moreover, the localization process of different stages by the proposed method with the sequence 00 of KITTI dataset are shown in Figure 5. We evaluated the localization accuracy by overlaying the localization trajectories against the reference trajectories, as shown in Figure 6. Note that the localization trajectories were achieved with a vision-only solution.

Figure 6 shows comparison of localization results by the proposed method with sequences 00, 05, 09 of the KITTI dataset. From which, it can be seen that the localization results of the proposed method are consistent with the reference trajectories, and in most of the regions, it performed reliably. Moreover, the localization accuracy does not depend on the accuracy at the last state of the vehicle, because even if there is a position deviation at the sharp turn, it can be corrected well later, as shown in the green circle in Figure 6a, b. Moreover, it performs well in the relatively flat loop, as shown in Figure 6c.

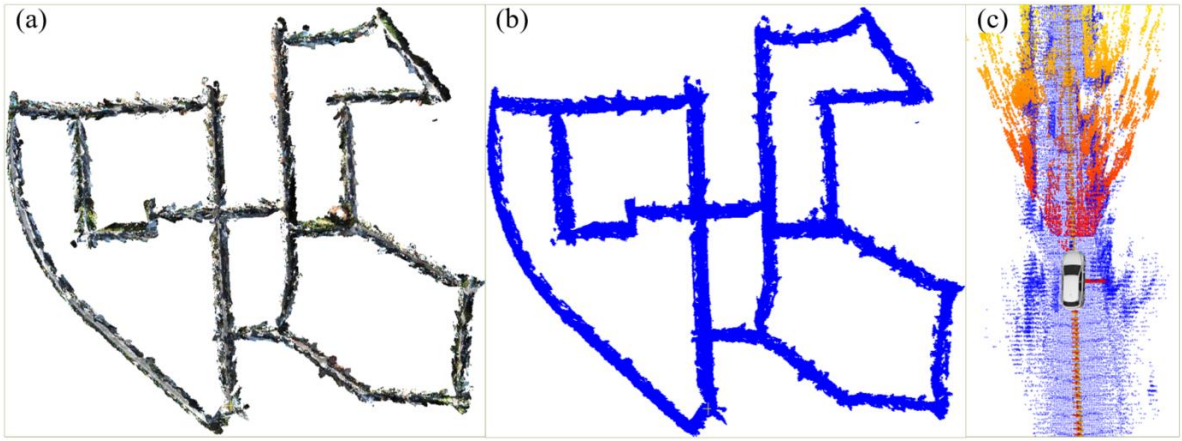

Figure 5. Localization process of different stages by the proposed method with the sequence 00 of the KITTI dataset. (**a**) Dense prior visual point cloud map; (**b**) priori visual point cloud map after downsampling; (**c**) local enlargement of localization result.

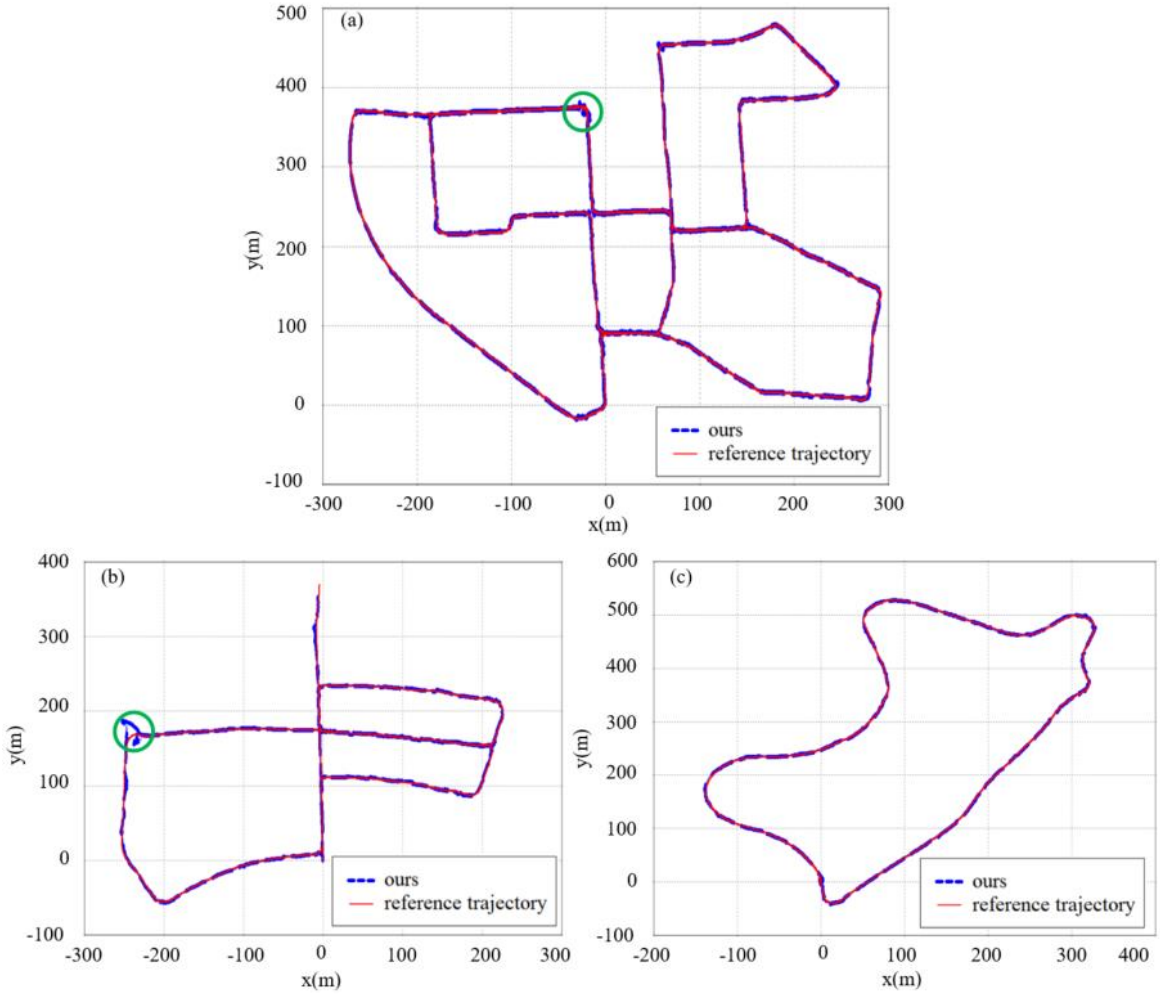

Figure 6. Comparison of localization results between the proposed method and the reference trajectories. (**a**) Sequence 00 of the KITTI dataset; (**b**) sequence 05 of KITTI dataset; (**c**) sequence 09 of KITTI dataset.

4.2.2. Evaluation with Field Test

The field test data Site-1 and Site-2 are used for experimental analysis. Moreover, the localization process of different stages by the proposed method with the field test data Site-2 are shown in Figure 7. We evaluated the localization accuracy by computing differences between the localization trajectories and the reference trajectories. As shown in Figure 8.

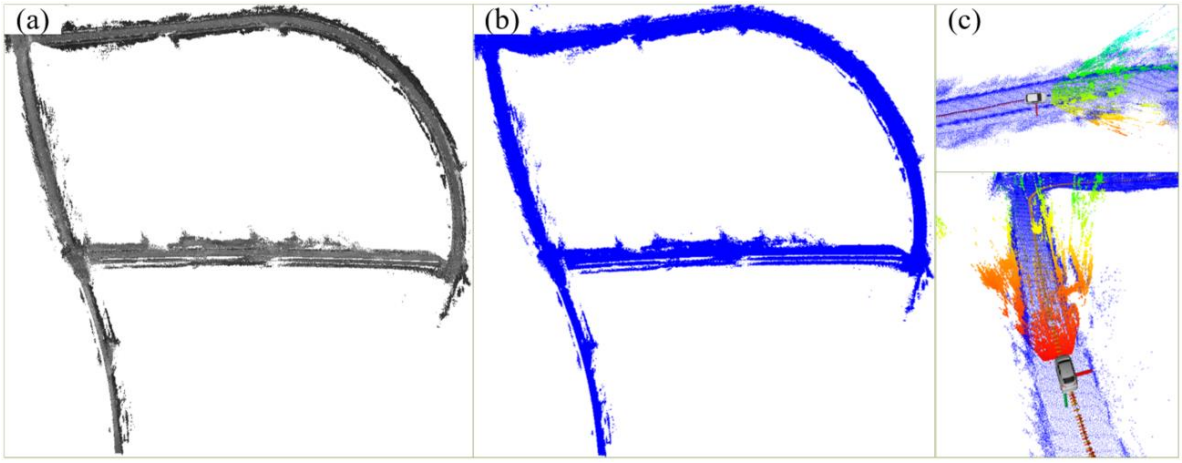

Figure 7. Localization process of different stages by the proposed method with the field test data Site-2. (**a**) Dense prior visual point cloud map; (**b**) priori visual point cloud map after downsampling; (**c**) local enlargement of localization result.

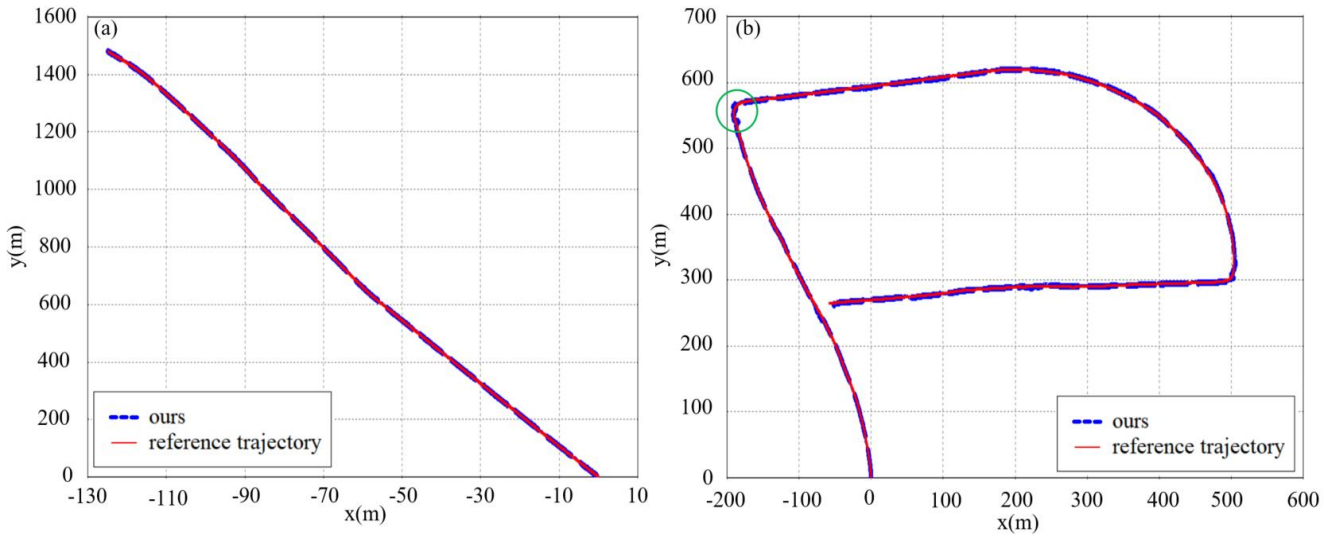

Figure 8. Comparison of localization results between the proposed method and the reference trajectories with field test data. (**a**) Field test data Site-1; (**b**) field test data Site-2.

Figure 8 shows comparison of localization results by the proposed method with the field test data Site-1 and Site-2. From the localization results of field test data in Figure 8, it can be seen that even though the localization trajectories were achieved with a vision-only solution, the consistency between the two trajectories is well preserved.

4.3. Quantitative Evaluation

To verify the proposed method, the RMSE of absolute localization trajectory translation and rotation are used to analyze the localization accuracy. Assume the rigid body transformation S corresponding to the least-squares solution that maps the localization

trajectory T_{loc} onto the reference trajectory T_{ref}. Then the RMSE of absolute localization trajectory translation and rotation can be computed as follows [43,44]:

$$t_{Error} = \left(\frac{1}{N}\sum_{i=1}^{N} \left\| trans\left(T_{ref,i}^{-1} S T_{loc,i}\right)\right\|^2\right)^{\frac{1}{2}} \tag{8}$$

$$R_{Error} = \left(\frac{1}{N}\sum_{i=1}^{N} \left\| \angle\left(R_{ref,i} R_{loc,i}^{\mathrm{T}}\right)\right\|^2\right)^{\frac{1}{2}} \tag{9}$$

where N is the number of localization points, $T_{ref,i}$ is the i-th reference trajectory point, $T_{loc,i}$ is the i-th localization trajectory point, $R_{ref,i}$ is the rotation matrix of the i-th reference point, $R_{loc,i}$ is the rotation matrix of the i-th localization point, $trans(\cdot)$ means take the translation vector of the variable inside the brackets, and $\angle(\cdot)$ represents converting the rotation matrix to Euler angles. The RMSE of absolute trajectory translation and rotation can describe the errors of the whole localization trajectory relative to the reference trajectory, as shown in Figure 9.

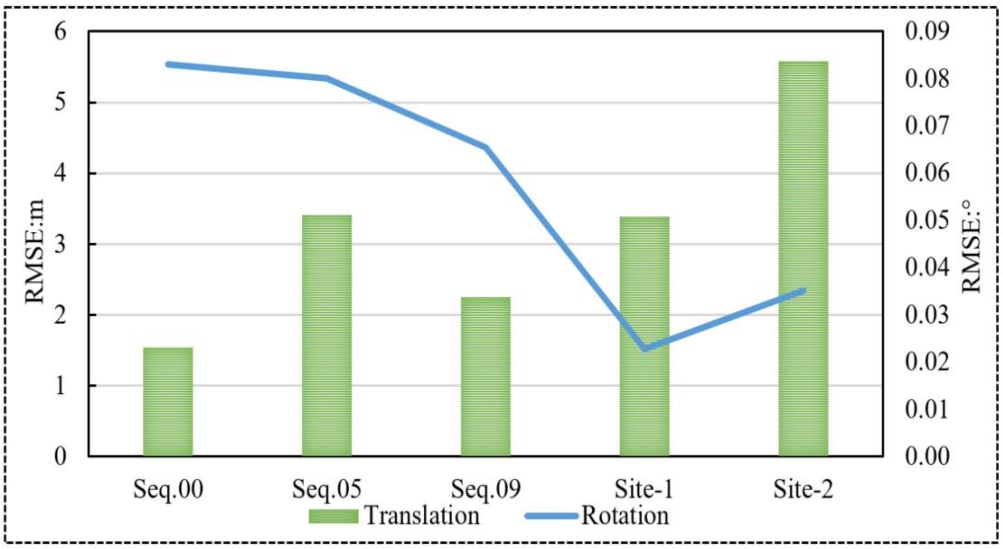

Figure 9. The root mean square error (RMSE) of absolute trajectory translation and rotation. Bar graph represents absolute translation error, line graph represents absolute rotation error.

From Figure 9, it can be seen that the RMSE of translation with sequences 00, 05, 09 of the KITTI dataset are less than 3.5 m, and the field test data Site-1 and Site-2 are 3.39 m and 5.59 m, respectively. The RMSE of rotation with sequences 00, 05, 09 of the KITTI dataset are less than 0.08°, and the field test data Site-1 and Site-2 are 0.02° and 0.04°, respectively.

4.4. Comparison with Other Methods

4.4.1. Comparison with Camera-Based Method

In order to confirm the localization results of the proposed method, we compared the localization performance to that of the state-of-the-art work by Kim et al. [45], which is a lightweight visual localization algorithm that performs localization within prior 3D LiDAR maps. The comparison of the proposed method and the work in [45] is shown in Table 1. We used average translation and rotation errors to evaluate the localization accuracy, which is achieved by averaging all of the differences between the localization trajectory and the reference trajectory.

Table 1. Comparison of average localization errors and standard deviation between the proposed method and the method in [45].

Scene	Ours				Kim et al. [45]			
	Translation/m		Rotation/°		Translation/m		Rotation/°	
	Average	Std.	Average	Std.	Average	Std.	Average	Std.
00	1.27	0.93	1.20	2.03	0.13	0.11	0.32	0.39
05	3.18	5.58	1.27	1.97	0.15	0.14	0.34	0.40
09	1.94	1.22	1.06	1.61	0.18	0.22	0.34	0.34

Table 1 summarizes localization errors with the KITTI dataset. In this table, the translation and rotation errors are averaged over all poses in each sequence and given in average localization errors and standard deviation. It can be seen that the average translation errors and standard deviation with the KITTI dataset by the proposed method are below 3.18 and 5.58 m. The average rotation errors and standard deviation with the KITTI dataset by the proposed method are under 1.27° and 2.03°. The method in [45] shows better accuracy with sequences 00, 05, 09 of the KITTI dataset. However, to minimize the depth residual of prior LIDAR depth and stereo depth, LiDAR depth is also used in the localization process. Note that our method updates the initial pose after matching between the current visual point cloud and the candidate sub-map, and the localization process is a vision-only solution.

4.4.2. Comparison with LiDAR-Based Method

We make further comparison with LiDAR-based localization [46]. Sequence 00 of the KITTI dataset is selected for this experiment. The comparison between the proposed method and the LiDAR-based method is shown in Figure 10. Where Figure 10a is the localization with the proposed method, Figure 10b is the localization with LiDAR-based method. Furthermore, it was almost equally divided into ten segments to calculate the horizontal localization errors relative to the reference trajectory, as shown in Figure 10c, the yellow five-pointed stars P1–P10 represent ten horizontal coordinate points, and the details of Figure 10c can be seen from local enlargement in Figure 10d. The corresponding comparison of horizontal segment localization errors between the proposed method and LiDAR-based method is shown in Table 2.

From Figure 10, it can be seen that the proposed method is consistent with the reference trajectory and LiDAR-based localization method both in global (as shown in Figure 10c) and local enlargement (as shown in Figure 10d). Moreover, local enlargement shows the details of the proposed method compared with reference trajectory and LiDAR-based method. As can be seen from Table 2, the horizontal segment localization errors between the proposed method and the LiDAR-based method are less than 3 m. The maximum error of the proposed method occurs at sharp turns P7, which is 2.96 m, and the average error is 1.12 m. The maximum error of LiDAR-based method occurs at P4, which is 1.73 m, and the average error is 0.93 m.

Figure 10. Comparison between the proposed method and LiDAR-based method with the sequence 00 of KITTI dataset. (**a**) Top view of the proposed method; (**b**) top view of LiDAR-based method; (**c**) comparison of horizontal segment localization results; (**d**) local enlargement of localization result.

Table 2. Comparison of horizontal segment localization errors between the proposed method and LiDAR-based method.

Methods	Segments Horizontal Localization Errors (m)										
	P1	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**	**P8**	**P9**	**P10**	**Average**
Ours	1.04	0.33	1.82	1.59	0.74	1.24	2.96	0.56	0.63	0.25	1.12
LiDAR-based method	1.60	0.25	1.45	1.73	1.38	0.95	0.87	0.35	0.32	0.44	0.93

4.5. Time Performance

Time performance in initial pose estimation (IPE), current visual point cloud generation (CVPCG), adaptive prior map segmentation (APMS), downsampling and data cleaning (DSDC), and localization are counted to verify the efficiency of the proposed method. The statistical results of each localization stage are shown in Table 3.

Table 3. Time performance of each localization process by the proposed method.

Sequences	Length (m)	Image Size	IPE (s)	CVPCG (s)	APMS (s)	DSDC (s)	Localization (s)	Average (s)
00	3723.30	1241 × 376	0.12	0.15	0.02	0.11	0.33	0.73
05	2203.91	1226 × 370	0.11	0.14	0.01	0.11	0.33	0.70
09	1704.76	1226 × 370	0.11	0.12	0.01	0.10	0.33	0.67
Site-1	1550.53	1280 × 640	0.08	0.23	0.01	0.21	0.34	0.87
Site-2	2248.42	1280 × 640	0.08	0.24	0.01	0.20	0.35	0.88

From Table 3, it can be seen that the time performance in initial pose estimation and adaptive prior map segmentation of the localization process is relatively small, while current visual point cloud generation, downsampling and data cleaning, and localization are slightly time-consuming. The average time consuming of sequences 00, 05, 09 of the KITTI dataset is less than 0.73 s per frame and the field test data is less than 0.88 s per frame, which verified the efficiency of the proposed method.

5. Discussion

To check the localization results of the proposed method, extensive experiments are performed to qualitatively and quantitatively evaluate the localization errors. From qualitative analysis (Figures 6 and 8), it can be seen that the localization results of the proposed method are consistent with the reference trajectories, and in most of the regions, it performed reliably. Moreover, the localization accuracy does not depend on the accuracy at the last state of the vehicle, because even if there is a position deviation at the sharp turn, it can be corrected well later, as shown in the green circle in Figures 6a,b and 8b. Moreover, it perform well in the relatively flat areas, as shown in Figures 6c and 8a. From quantitative evaluation (Figure 9), it can be seen that the proposed method achieved good localization results. Although the translation accuracy of field test data Site-1 and Site-2 are lower than that of KITTI dataset, it shows the feasibility of the proposed method in a real world. We would like to point out that our main purpose was to verify a vision-only solution within GNSS-challenged environments. From Figure 10 and Table 2, it can be seen that the proposed method can achieve similar localization accuracy compared to LiDAR-based method. However, there are still some problems, such as the localization results at the sharp turns are not smooth enough; thus, the localization accuracy needs to be further improved.

The segmentation threshold is related to operation efficiency and the density of prior point cloud map. If the segmentation threshold is too large, the efficiency of localization is relatively low; if the segmentation threshold is too small, the segmentation is more time-consuming, and the localization is efficient. Therefore, it is necessary to determine a suitable threshold. In this paper, the threshold is set to 80 m. When the point cloud is dense, the segmentation threshold can be appropriately reduced, and when the point cloud is sparse, the segmentation threshold can be appropriately enlarged to ensure the efficiency of localization.

6. Conclusions

In this paper—which is different from camera localization methods based on prior LiDAR maps—we proposed a novel localization method for an autonomous vehicle with prior visual point cloud constraints generated by a stereo camera, which cannot only reduce the localization cost, but also improve accuracy and stability. The proposed method is conducted by the SGBM algorithm for estimating the visual point cloud of each image, and stereo visual odometry for providing the initial position. Then, multiple filtering and adaptive priori map segmentation are introduced for fast matching and localization; and the candidate sub-map is matched with the visual point cloud of each frame by NDT to update the pose prediction based on last frame for accurate localization, which overcomes the error accumulation due to long-running without loop optimization. Comprehensive

experiments show that it is capable of autonomous vehicle localization with prior visual point cloud constraints in GNSS-challenged environments.

As the proposed localization method highly depends on the prior visual point cloud map, if there are major changes in the environment, such as road reconstruction, architecture construction, or demolition, the localization accuracy may decrease or even fail. Therefore, methods to update the prior maps quickly and efficiently with low-cost sensors will be our future work. Moreover, we will enhance localization accuracy by integrating GNSS/INS assisted constraints when the GNSS signal is reliable.

Author Contributions: Conceptualization, X.L., F.W. and B.Y.; data curation, W.Z.; formal analysis, X.L., F.W. and B.Y.; funding acquisition, F.W. and B.Y.; methodology, X.L.; project administration, F.W.; supervision, F.W. and B.Y.; validation, X.L. and W.Z.; visualization, X.L.; writing—original draft, X.L.; writing—review and editing, F.W. and B.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China for Distinguished Young Scholars (grant no. 41725005), the Key Project of the National Natural Science Foundation of China (grant no. 41531177) and the National Key Research and Development Program of China (grant no. 2016YFB0501803).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Wen, W.; Hsu, L.-T.; Zhang, G. Performance Analysis of NDT-based Graph SLAM for Autonomous Vehicle in Diverse Typical Driving Scenarios of Hong Kong. *Sensors* **2018**, *18*, 3928. [CrossRef] [PubMed]
2. Carvalho, H.; Del Moral, P.; Monin, A.; Salut, G. Optimal nonlinear filtering in GPS/INS integration. *IEEE Trans. Aerospace Electron. Syst.* **1997**, *33*, 835–850. [CrossRef]
3. Mohamed, A.H.; Schwarz, K.P. Adaptive kalman filtering for INS/GPS. *J. Geodesy* **1999**, *73*, 193–203. [CrossRef]
4. Wang, D.; Xu, X.; Zhu, Y. A Novel Hybrid of a Fading Filter and an Extreme Learning Machine for GPS/INS during GPS Outages. *Sensors* **2018**, *18*, 3863. [CrossRef] [PubMed]
5. Liu, H.; Ye, Q.; Wang, H.; Chen, L.; Yang, J. A Precise and Robust Segmentation-Based Lidar Localization System for Automated Urban Driving. *Remote Sens.* **2019**, *11*, 1348. [CrossRef]
6. Nuchter, A.; Lingemann, K.; Hertzberg, J.; Surmann, H. 6D SLAM-3D mapping outdoor environments. *J. Field Robot.* **2007**, *24*, 699–722. [CrossRef]
7. Bosse, M.; Zlot, R.; Flick, P. Zebedee: Design of a spring-mounted 3-d range sensor with application to mobile mapping. *IEEE Trans. Robot.* **2012**, *28*, 1104–1119. [CrossRef]
8. Suzuki, T.; Kitamura, M.; Amano, Y.; Hashizume, T. 6-DOF localization for a mobile robot using outdoor 3D voxel maps. In Proceedings of the 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Taipei, Taiwan, 18–22 October 2010; pp. 5737–5743.
9. Yoneda, K.; Tehrani, H.; Ogawa, T.; Hukuyama, N.; Mita, S. Lidar scan feature for localization with highly precise 3-D map. In Proceedings of the 2014 IEEE Intelligent Vehicles Symposium, Dearborn, MI, USA, 8–11 June 2014; pp. 1345–1350.
10. Ruchti, P.; Steder, B.; Ruhnke, M.; Burgard, W. Localization on openstreetmap data using a 3D laser scanner. In Proceedings of the 2016 IEEE International Conference on Robotics and Automation (ICRA), Seattle, WA, USA, 26–30 May 2015; pp. 5260–5265.
11. Stewart, A.D.; Newman, P. LAPS-localisation using appearance of prior structure: 6-dof monocular camera localisation using prior pointclouds. In Proceedings of the 2012 IEEE International Conference on Robotics and Automation (ICRA), Saint Paul, MN, USA, 14–18 May 2012; pp. 2625–2632.
12. Wolcott, R.W.; Eustice, R.M. Visual localization within LIDAR maps for automated urban driving. In Proceedings of the 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Chicago, IL, USA, 14–18 September 2014; pp. 176–183.
13. Li, Q.; Zhu, J.S.; Liu, J.; Cao, R.; Fu, H.; Garibaldi, J.M.; Li, Q.Q.; Liu, B.Z.; Qiu, G.P. 3D map-guided single indoor image localization refinement. *ISPRS J. Photogramm. Remote Sens.* **2020**, *161*, 13–26. [CrossRef]
14. Neubert, P.; Schubert, S.; Protzel, P. Sampling-based methods for visual navigation in 3D maps by synthesizing depth images. In Proceedings of the 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Vancouver, BC, Canada, 24–28 September 2017; pp. 2492–2498.

15. Xu, Y.Q.; John, V.; Mita, S.; Tehrani, H.; Ishimaru, K.; Nishino, S. 3D point cloud map based vehicle localization using stereo camera. In Proceedings of the 2017 IEEE Intelligent Vehicles Symposium (IV), Los Angeles, CA, USA, 11–14 June 2017; pp. 487–492.
16. Geiger, A.; Lenz, P.; Urtasun, R. Are we ready for autonomous driving? The KITTI vision benchmark suite. In Proceedings of the 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Providence, RI, USA, 16–21 June 2012; pp. 3354–3361.
17. Choi, J. Hybrid map-based SLAM using a Velodyne laser scanner. In Proceedings of the 17th International IEEE Conference on Intelligent Transportation Systems (ITSC), Qingdao, China, 8–11 October 2014; pp. 3082–3087.
18. Im, J.H.; Im, S.-H.; Jee, G.I. Vertical corner feature based precise vehicle localization using 3D LIDAR in urban area. *Sensors* **2016**, *16*, 1268. [CrossRef] [PubMed]
19. Levinson, J.; Montemerlo, M.; Thrun, S. Map-based precision vehicle localization in urban environments. In Proceedings of the Robotics: Science and Systems III, Atlanta, GA, USA, 27–30 June 2007; p. 1.
20. Levinson, J.; Thrun, S. Robust vehicle localization in urban environments using probabilistic maps. In Proceedings of the 2010 IEEE International Conference on Robotics and Automation (ICRA), Anchorage, AK, USA, 3–7 May 2010; pp. 4372–4378.
21. Kim, D.; Chung, T.; Yi, K. Lane map building and localization for automated driving using 2D laser rangefinder. In Proceedings of the 2015 IEEE Intelligent Vehicles Symposium (IV), Seoul, Korea, 28 June–1 July 2015; pp. 680–685.
22. Qin, B.; Chong, Z.J.; Bandyopadhyay, T.; Ang, M.H.; Frazzoli, E.; Rus, D. Curb-intersection feature based Monte Carlo Localization on urban roads. In Proceedings of the 2012 IEEE International Conference on Robotics and Automation (ICRA), Saint Paul, MN, USA, 14–18 May 2012; pp. 2640–2646.
23. Forster, C.; Pizzoli, M.; Scaramuzza, D. Air-ground localization and map augmentation using monocular dense reconstruction. In Proceedings of the 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Tokyo, Japan, 3–7 November 2013; pp. 3971–3978.
24. Steder, B.; Ruhnke, M.; Burgard, W. Monocular camera localization in 3D lidar maps. In Proceedings of the 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Daejeon, Korea, 9–14 October 2016; pp. 1926–1931.
25. Brubaker, M.A.; Geiger, A.; Urtasun, R. Map-based probabilistic visual self-localization. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 652–665. [CrossRef] [PubMed]
26. Ziegler, J.; Lategahn, H.; Schreiber, M.; Keller, C.G.; Knöppel, C.; Hipp, J.; Haueis, M.; Stiller, C. Video based localization for bertha. In Proceedings of the 2014 IEEE Intelligent Vehicles Symposium Proceedings, Dearborn, MI, USA, 8–11 June 2014; pp. 1231–1238.
27. Radwan, N.; Tipaldi, G.D.; Spinello, L.; Burgard, W. Do you see the bakery? Leveraging geo-referenced texts for global localization in public maps. In Proceedings of the 2016 IEEE international conference on robotics and automation (ICRA), Stockholm, Sweden, 16–21 May 2016; pp. 4837–4842.
28. Spangenberg, R.; Goehring, D.; Rojas, R. Pole-based localization for autonomous vehicles in urban scenarios. In Proceedings of the 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Daejeon, Korea, 9–14 October 2016; pp. 2161–2166.
29. Lyrio, L.J.; Oliveira-Santos, T.; Badue, C.; De Souza, A.F. Image-based mapping, global localization and position tracking using vg-ram weightless neural networks. In Proceedings of the 2015 IEEE International Conference on Robotics and Automation (ICRA), Seattle, WA, USA, 26–30 May 2015; pp. 3603–3610.
30. Oliveira, G.L.; Radwan, N.; Burgard, W.; Brox, T. Topometric localization with deep learning. *arXiv* **2017**, arXiv:1706.08775.
31. Lin, X.H.; Yang, B.S.; Wang, F.H.; Li, J.P.; Wang, X.Q. Dense 3D surface reconstruction of large-scale streetscape from vehicle-borne imagery and LiDAR. *Int. J. Digit. Earth* **2020**, 1–21. [CrossRef]
32. Chang, J.R.; Chen, Y.S. Pyramid stereo matching network. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 5410–5418.
33. Lin, X.H.; Wang, F.H.; Guo, L.; Zhang, W.W. An automatic key-frame selection method for monocular visual odometry of ground vehicle. *IEEE Access* **2019**, *7*, 70742–70754. [CrossRef]
34. Rusu, R.B.; Cousins, S. 3D is here: Point cloud library (PCL). In Proceedings of the 2011 IEEE International Conference on Robotics and Automation (ICRA), Shanghai, China, 9–13 May 2011; pp. 1–4.
35. Kim, H.; Liu, B.; Goh, C.Y.; Lee, S.; Myung, H. Robust Vehicle Localization Using Entropy-Weighted Particle Filter-based Data Fusion of Vertical and Road Intensity Information for a Large Scale Urban Area. *IEEE Robot. Autom. Lett.* **2017**, *2*, 1518–1524. [CrossRef]
36. Hirschmuller, H. Stereo Processing by Semiglobal Matching and Mutual Information. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 328–341. [CrossRef] [PubMed]
37. Scaramuzza, D.; Fraundorfer, F. Visual odometry: Part I, the first 30 years and fundamentals. *IEEE Robotics and Automation Magazine.* **2011**, *18*, 80–91. [CrossRef]
38. Mur-Artal, R.; Tardós, J.D. Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras. *IEEE Trans. Robot.* **2017**, *33*, 1255–1262. [CrossRef]
39. Kuemmerle, R.; Grisetti, G.; Strasdat, H.; Konolige, K.; Burgard, W. g2o: A general framework for graph optimization. In Proceedings of the 2011 IEEE international conference on robotics and automation (ICRA), Shanghai, China, 9–13 May 2011; pp. 3607–3613.

40. Huhle, B.; Magnusson, M.; Straßer, W.; Lilienthal, A.J. Registration of colored 3d point clouds with a kernel-based extension to the normal distributions transform. In Proceedings of the 2008 IEEE international conference on robotics and automation (ICRA), Pasadena, CA, USA, 19–23 May 2008; pp. 4025–4030.
41. Magnusson, M.; Nuchter, A.; Lorken, C.; Lilienthal, A.J.; Hertzberg, J. Evaluation of 3d registration reliability and speed-a comparison of ICP and NDT. In Proceedings of the 2009 IEEE international conference on robotics and automation (ICRA), Kobe, Japan, 12–17 May 2009; pp. 3907–3912.
42. Magnusson, M. The Three-Dimensional Normal-Distributions Transform: An Efficient Representation for Registration, Surface Analysis, and Loop Detection. Ph.D. Thesis, Örebro University, Örebro, Sweden, 2013.
43. Zhang, Z.; Scaramuzza, D. A Tutorial on Quantitative Trajectory Evaluation for Visual (-Inertial) Odometry. In Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, Spain, 1–5 October 2018; pp. 7244–7251.
44. Sturm, J.; Engelhard, N.; Endres, F.; Burgard, W.; Cremers, D. A benchmark for the evaluation of RGB-D SLAM systems. In Proceedings of the 2012 IEEE/RSJ international conference on intelligent robots and systems (IROS), Vilamoura, Portugal, 7–12 October 2012; pp. 573–580.
45. Kim, Y.J.; Jeong, J.Y.; Kim, A. Stereo Camera Localization in 3D LiDAR Maps. In Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, Spain, 1–5 October 2018; pp. 1–9.
46. Shan, T.; Englot, B. LeGO-LOAM: Lightweight and Ground-Optimized Lidar Odometry and Mapping on Variable Terrain. In Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, Spain, 1–5 October 2018; pp. 4758–4765.