remote sensing

MDPI

Article

Indoor Environment Reconstruction for Unmanned System Using Multiple Low-Cost Sensors

Yunshu Wang [1], Bin Ding [1], Haiqing Wang [1], Qian Meng [2,3,*], Yuan Zhuang [4] and Haonan Jia [5]

[1] Nanjing Research Institute of Electronic Engineering, Nanjing 210007, China; wangys@nuaa.edu.cn (Y.W.); binding_nj@163.com (B.D.); hqwang_28@163.com (H.W.)
[2] School of Instrument Science and Engineering, Southeast University, Nanjing 210096, China
[3] Key Laboratory of Micro-Inertial Instruments and Advanced Navigation Technology, Ministry of Education, Southeast University, Nanjing 210096, China
[4] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430072, China; yuan.zhuang@whu.edu.cn
[5] State Key Laboratory of Satellite Navigation System and Equipment Technology, The 54th Research Institute of China Electronics Technology Group Corporation, Shijiazhuang 050002, China; 230219409@seu.edu.cn
* Correspondence: qianmeng@seu.edu.cn

Abstract: Considering that laser scan stations are expensive and heavy for the indoor environment reconstruction of unmanned systems, a low-cost and light reconstruction system is proposed in this research. The system consists of a cross-structured light visual (SLV) sensor and an inertial navigation system (INS). The cross-SLV sensor is used to scan the surroundings and to estimate the angle change between two adjacent measurements. To improve the robustness and accuracy of the angle measurement, a Kalman Filter (KF) with simple construction is designed to fuse the inertial information from the INS. The factors which influence ranging accuracy are analyzed and ranging experiments show that the SLV sensor has an accuracy of higher than 2% when the distance is less than 4 m. Then the reconstruction results of a kitchen and corridor show that the error of most points is less than 50 mm for either kitchen (94%) or corridor (85%), and the mean errors and standard deviations of kitchen and corridor are less than 20 mm and 30 mm, respectively.

Keywords: reconstruction; low cost; structured light sensor; visual sensor; inertial sensor

check for updates

Citation: Wang, Y.; Ding, B.; Wang, H.; Meng, Q.; Zhuang, Y.; Jia, H. Indoor Environment Reconstruction for Unmanned System Using Multiple Low-Cost Sensors. *Remote Sens.* **2024**, *16*, 899. https://doi.org/10.3390/rs16050899

Academic Editor: Pablo Rodríguez-Gonzálvez

Received: 20 January 2024
Revised: 29 February 2024
Accepted: 29 February 2024
Published: 3 March 2024

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

1. Introduction

Three-dimensional reconstruction is well established and accepted in manufacturing, automobile and construction industries, and there is keen interest shown in investigating how this can be applied in practice [1]. One of the most interesting fields is the environment reconstruction of unmanned systems such as Micro Aerial Vehicles (MAVs) in unknown environments [2]. Environment reconstruction is significant for the navigation of MAVs in indoor environments [3].

Three-dimensional environment reconstruction can be divided into two types: laser scanning-based reconstruction and vision-based reconstruction. The former uses the depth distance of scenes acquired by range sensors and transforms the measured points. The latter searches the feature points of the environment using a vision-based system [4,5].

In previous decades, a laser scanning technique was developed to build a 3D environment model with time efficiency and high accuracy [6,7]. Laser scanners measure the distance, horizontal angle and vertical angle of a point to obtain the 3D coordinate of the point. For a commercial laser scanner, the scanning module is mounted inside with high frequency so that it can measure a full scene with millimeter-to-centimeter accuracy at speeds of thousands to hundreds of thousands of measurements per second [8]. Due to the high accuracy of terrestrial laser scanning (TLS) in both distance measurement and angle measurement, the most widely researched aspect of laser scanning-based reconstruction is

how to use points clouds of environments rather than how to obtain the points clouds. The points cloud obtained from a laser scan station is used to build detailed models of walls, doors, windows, etc., which are called building information models (BIMs) [9,10]. However, the high cost and weight of equipment and the difficulty of cloud point processing must still be improved.

Several vision-based systems have been proposed for 3D environment reconstruction due to their low hardware cost. The general process can be broken down into the following steps: feature points extraction, image matching, dense point cloud achievement, surface reconstruction and textured mapping [11]. Nevertheless, vision-based systems are limited to reconstructing an environment with low illumination and sparse or repetitive features, due to the problem of calculating depth by finding correspondence between two images [12].

Structured light visual (SLV) sensors, which consist of a structured light projector and a monocular camera or binocular camera, can calculate the depth of a light strip in an image by the laser triangulation algorithm. This is widely used in measurement areas such as automatic agriculture [13], industry inspection [14,15] and robot indoor navigation [16,17], due to the advantages of simple construction, high accuracy and low cost. An SLV sensor mounted on an MAV can scan the environment agilely with rotation of the MAV. Thus, the registration of the points cloud between two adjacent measurements is a key problem for 3D environment reconstruction of MAVs with an SLV sensor.

In this paper, which is aimed at addressing the challenges that laser scanning-based reconstruction is too expensive and too heavy for MAVs and that vision-based reconstruction is invalid in environments with low illumination and sparse or repetitive features, a low-cost and light 3D environment reconstruction system based on multi-sensors is proposed. The system consists of a monocular camera, a cross-structured light and an inertial navigation system (INS). The cross-structured light, which consists of a vertical structured light and a horizontal structured light, is calibrated first. The vertical structured light is used to scan the surroundings while the horizontal structured light is used to estimate the heading angle change between adjacent images. To improve the robustness and accuracy of the heading angle estimation, a Kalman Filter (KF) is added to the system to fuse the measurements of the SLV sensor and INS. The experimental results show that the distance measured by the low-cost system has centimeter accuracy at a distance of about 4 m in indoor environments, and different kinds of indoor environments reconstructed by the proposed system are accurate enough for navigation of MAVs.

2. Literature Review

An important issue for 3D environment reconstruction is whether the laser scanner or the vision sensor can only acquire points in their field of view. In order to acquire full data from a scene, multiple measurements from different viewpoints have to be recorded. Then they have to be registered accurately in a common coordinate system [18].

2.1. Features-Based Registration

Features-based registration is widely used for vision-based systems. To register feature points such as SIFT [19], FAST [20] and ORB [21] from different viewpoints, Structure from Motion (SfM) is widely researched for the reconstruction of vision-based systems. SfM can estimate the motion of a camera and transform the 2D coordinates of points in an image frame to the 3D coordinates of points in a world frame. The most widely available SfM solution is Bundler [22]. In recent years, various improvements in SfM have been proposed to increase the robustness and accuracy. In [23], a framework of 3D reconstruction on an inexpensive robotic platform using a webcam and robot wheel odometry is proposed. The results show that the technique is efficient and robust to a variety of environment scenarios with different scales and sizes. In [24], a novel optical flow-based SfM is proposed for the reconstruction of surfaces with few textures using video sequences acquired under strong illumination changes. In [25], an advanced energy function for salient image features is established by combining multiple feature types with a weighted finite-element mesh

to prolong the tracking lifetimes of consecutive frames and ensure the spatiotemporal consistency and robustness of homonymous image features in different subsequences. However, when an environment is lacking illumination, or the features have repetitive patterns, the performance of features-based registration drops rapidly.

2.2. Points Cloud-Based Registration

A deep camera or Time-of-Flight (ToF) camera can solve the problem above. These cameras can obtain the points cloud of an image. The landmark contribution of points cloud registration is the Iterative Closest Point (ICP) method. Many researchers have devoted great efforts to improving ICP, especially its speed and robustness. In [26], to improve the robustness of ICP, an inequality constraint of a rotation angle is introduced into the registration model, which is solved by an extended ICP algorithm. A 3D LiDAR point cloud data registration method based on the ICP algorithm, which is improved by the Gaussian mixture model (GMM) considering corner features, is proposed in [27] to address timeliness and local optima. In [28], a hybrid approach which combines ICP-based registration with feature extraction and surface reconstruction is proposed to deal explicitly with the inhomogeneous density of point clouds produced by LiDAR scanners. Even though such improvements in ICP are proposed, the cost of calculation is still enormous.

2.3. Distance- and Angle-Based Registration

Although the ICP is used to register the points cloud generated by a 3D laser scanner, most 3D laser scanners use distance- and angle-based registration to obtain surrounding points clouds. The laser scanner measures the distances in a line and rotates to scan the environment. Due to the high accuracy in measuring distance and angle, the generated points cloud is accurate too. In this case, the cost of 3D laser scanners is usually expensive. Some researchers try to use other low-cost equipment to realize a reconstruction. In [29], a rangefinder which can measure distance, horizontal angle and vertical angle is considered as a laser scan station to build indoor 3D environments. The rangefinder is also used monitor 3D snow accumulation and melt dynamics with a pan-tilt unit [30,31].

2.4. Low-Cost 3D Environment Reconstruction System

Regarding a low-cost 3D environment reconstruction system, an SLV sensor is the optimal solution. In contrast to the application of industry measurement, which requires the measured distance to have millimeter accuracy and only works in suitable conditions [32], an SLV sensor for environment reconstruction requires lower accuracy (centimeters) in ranging but needs to deal with challenging conditions such as a varying scanning rate and a change in illumination. Even though an SLV sensor can obtain the depth of a partial image by one measurement, the field of view of the structured light sensor needs to be increased for environment reconstruction. A common method is using an omnidirectional vision system [33] or mounting the SLV sensor onto a rotated module [34], but this hardware is too complex for application to MAV [35].

With the development of artificial intelligence (AI) and computing chips, some novel methods employing neural networks or deep learning can significantly simplify the hardware architecture of 3D reconstruction systems. A Neural Radiance Field (NeRF), along with a series of improvement works, utilizes neural networks to reconstruct the shape and corresponding texture of targets or environments in three dimensions with only a series of consecutive images [36–38]. A modular framework for modularized NeRF development has been proposed to forming a plug-and-play architecture for multi-sensor information fusion [39]. Neural networks are also applied in multi-source information fusion of 3D reconstruction, with approaches such as combining CNN with the KF to predict a noise model [40] or correction phase [41]. There are also studies proposing factor graph-based estimation for learning tactile models and ground encoding using CNN [42,43]. In order to simplify the software architecture and reduce computational requirements, a novel learnable observation noise model for robust pose estimation by integrating type-2 TS FIS

and factor graph optimization was proposed [44], which uses a visual LiDAR inertial sensor. However, these machine learning methods still require high-performance computing units to provide substantial computational support during both training and real-time calculation processes, which is not the original intention of this study.

2.5. Main Contributions

To sum up, laser scanning-based reconstruction is too expensive and too heavy for unmanned systems. Vision-based reconstruction is light enough but usually invalid in an environment with low illumination and sparse or repetitive features. Other low-cost SLV systems with omnidirectional vision systems or rotated modules are too complex for application to unmanned systems. In contrast with previous works, we highlight our contributions as follows:

(1) Taking into consideration that laser scan stations are expensive and heavy for indoor environment reconstruction, a low-cost environment reconstruction system which consists of an SLV sensor and INS is proposed. This is much cheaper and lighter than the laser scan station that laser scanning-based reconstruction needs. It is also lighter than the other low-cost environment reconstruction systems using omnidirectional vision systems or rotated modules.
(2) To improve the robustness and accuracy of the angle measurement, a Kalman Filter (KF) with simple construction is designed to fuse the inertial information from INS. This is significant because this reconstruction system can work in an environment with low illumination and sparse features, which will make vision-based reconstruction invalid.
(3) An experiment of environment reconstruction is proceeded with. The reconstruction results of two kinds of environment show that both the main construction and some details can be reflected clearly. The reconstruction results of the kitchen and corridor show that the error of most points is less than 50 mm for either kitchen (94%) or corridor (85%) and that the mean error and standard deviation of kitchen and corridor are less than 20 mm and 30 mm, respectively.

3. Methodology

In this part, the coordinates of the points on structured light planes are calculated first. The vertical structured light plane is used to scan the surroundings and the horizontal structured light plane is used to estimate the angle change between two adjacent measurements. Then, a KF with simple construction is designed to fuse the inertial information from INS. The framework of methodology is shown in Figure 1. In particular, the core algorithms are summarized in the red dashed box. After collecting an image containing cross-structured light, calibration of the SLV sensor is performed (the details will be described in Section 3.2) to determine the rotation and translation relationship between the cross-structured light sensor and the camera. The three-dimensional coordinates of the structured light in the camera frame $\{C\}$ can then be calculated. By solving for the Euclidean distance, the distance from any structured light spot to the camera in each frame can be obtained. During the process of rotating and scanning the environment in the 3D reconstruction system, the angle between two frames is calculated through feature-based image registration to compensate for low illumination or a lack of features leading to registration failure. In cases where image registration fails, the angle between two frames is calculated using the method described in Section 3.3.1. Subsequently, the KF method described in Section 3.3.2 is used to fuse the angle measurements from the inertial sensor and the SLV sensor, yielding more accurate inter-frame angles. These inter-frame angles represent the rotation relationship of the camera frame $\{C\}$ at each adjacent moment. Finally, by mapping the measured structured light coordinates in each frame to the corresponding moment's camera frame $\{C\}$ using the inter-frame angles, a three-dimensional environment reconstructed by the SLV sensor scanning a full circle is obtained.

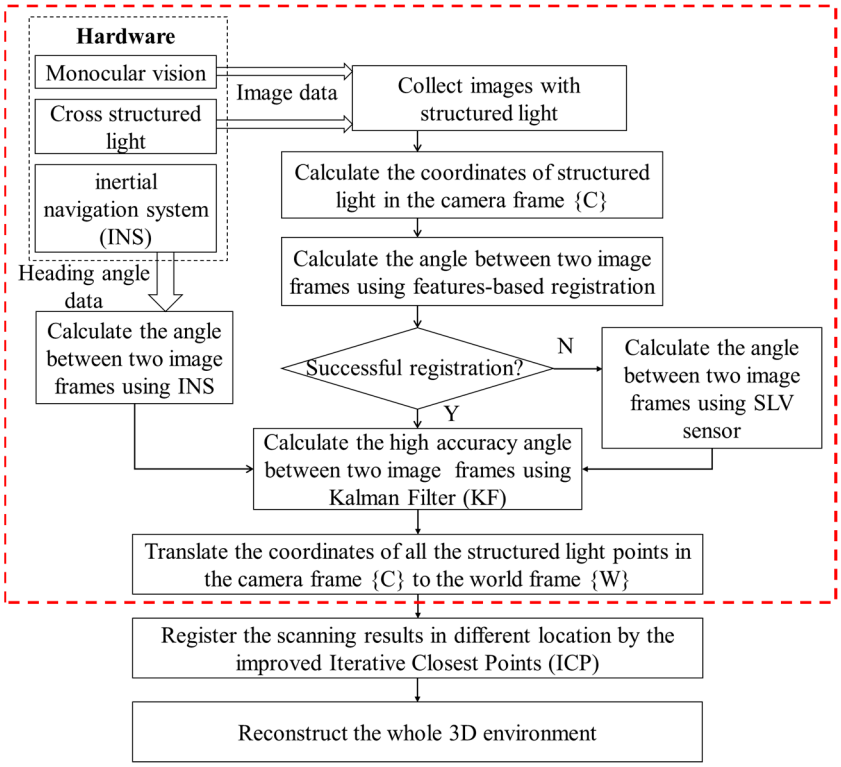


Figure 1. Overview of the methodology.

3.1. Definitions of Coordinates

When the reconstruction system works, four reference frames are considered:

The world frame {*W*}: The points cloud generated by the SLV sensor is modelled in this frame. In this paper, the *x*, *y*, *z*-axes are aligned with the east, north and vertical axes, respectively.

The camera frame {*C*}: This frame is attached to the monocular camera, with its origin at the optical center of the camera and with the z-axis pointing along the optical axis.

The image frame {*o*}: This frame is a two-dimensional frame, with its origin at the top-left of the image that is captured by the camera.

The cross-structured light frame {*L*}: This frame is attached to the cross-structured light projector. Its origin is the fire point of the projector. The *z*-axis is along the intersecting line of two light planes. The *x*-axis and *y*-axis are perpendicular to the *z*-axis in each structured light plane.

The camera projection frame {*C'*}: This frame is on the ground. Its origin is the projection of the camera frame's origin on the ground, and the *x*-axis and *z*-axis are the projections of the camera frame's *x*-axis and *z*-axis, respectively. The *y*-axis is perpendicular to the ground and points down.

The INS frame {*I*}: This is the frame of INS, with its origin at the center of the INS body. And the *x*, *y*, *z*-axes denote the front, right and down directions of the INS body, respectively.

The relationship of these frames is shown in Figure 1. The relative pose between frame {*I*} and frame {*C*} has been calibrated and is constant, while the relative pose between frame {*I*} and frame {*W*} is variable. The main purpose of this paper is calculating the relative pose between frame {*I*} and frame {*W*}. In this case, the initial pose of frame {*I*} is considered as the origin of frame {*W*}.

3.2. Distance Measurement by the SLV Sensor

Assume that P is an arbitrary point on the structured light plane. Point p is the mapping of P on the image frame. The coordinates of P in the camera frame are (x_P^c, y_P^c, z_P^c).

The coordinates of image point p in the image frame are (u_p, v_p). According to [45], the relationship of $\{C\}$ and $\{o\}$ can be represented as:

$$z_p^c \begin{bmatrix} u_p \\ v_p \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{f}{dx} & 0 & u_0 \\ 0 & \frac{f}{dy} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_P^c \\ y_P^c \\ z_P^c \end{bmatrix} = \mathrm{K} \begin{bmatrix} x_P^c \\ y_P^c \\ z_P^c \end{bmatrix} \tag{1}$$

where K is the internal parameters matrix of the camera. This can be obtained through the camera calibration 394../*.toolbox for MATLAB [46]. The dx and dy are the effective pixel pitches in directions x and y on the image plane, respectively; f is the focal length of the camera; and (u_0, v_0) are the coordinates of the optical center on the image frame.

The relationship between $\{L\}$ and $\{C\}$ can be represented as [17]:

$$\begin{bmatrix} x_p^c \\ y_p^c \\ z_p^c \\ 1 \end{bmatrix} = \begin{bmatrix} {}_L^C\mathrm{R} & {}_L^C\mathrm{T} \\ 0_{1\times 3} & 1 \end{bmatrix} \begin{bmatrix} x_p^l \\ y_p^l \\ z_p^l \\ 1 \end{bmatrix} \tag{2}$$

where the rotation matrix is ${}_L^C\mathrm{R} = \begin{bmatrix} {}_L^Cr_{11} & {}_L^Cr_{12} & {}_L^Cr_{13} \\ {}_L^Cr_{21} & {}_L^Cr_{22} & {}_L^Cr_{23} \\ {}_L^Cr_{31} & {}_L^Cr_{32} & {}_L^Cr_{33} \end{bmatrix}$ and the translation matrix is ${}_L^C\mathrm{T} = \begin{bmatrix} {}_L^Ct_1 & {}_L^Ct_2 & {}_L^Ct_3 \end{bmatrix}^T$, which means the distance from vertical structured light plane to the camera center is ${}_L^Ct_1$ and the distance from the horizontal structured light plane to the camera center is ${}_L^Ct_2$. Both can be estimated by calibration as follows [47,48]:

(1) Obtain the center points of the structured light on the calibration plate by the gray gravity center method.
(2) Fit the center points of the structured light by the least square method to obtain the linear equation of structured light on the image frame $\{o\}$.
(3) Fix the SLV sensor and change the pose of the calibration plate; repeat steps (1)–(2) to capture k images and the correlative linear equations of structured light on the image frame.
(4) Plane fitting. The equation of the structured light plane in the camera frame can be obtained by plane fitting. In theory, two structured light line equations can determine a structured light plane. However, due to the error in extracting the center of the structured light line, it cannot be guaranteed that the two structured light lines will be in the same plane. Therefore, it is necessary to take more than three images and perform plane fitting (as shown in Figure 2).

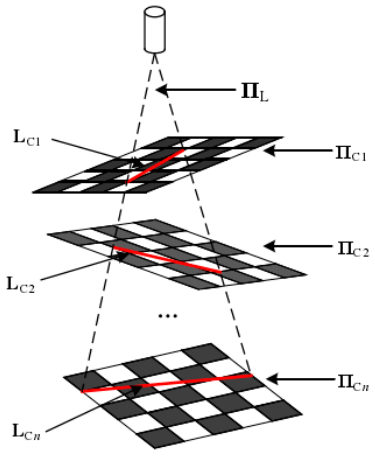
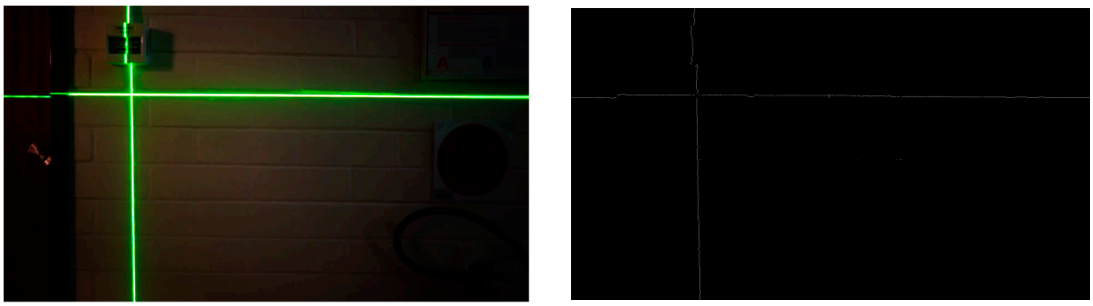


Figure 2. Schematic diagram of structured light plane fitting.

According to Equations (1) and (2), all the points on structured light planes can be calculated. First of all, the light strips on the image frame need to be detected, as shown in Figure 3.



(a) **(b)**

Figure 3. The detection of cross-structured light. (**a**) The original image of cross-structured light; (**b**) the image after light strips detection.

For an arbitrary point p on the image frame, the coordinates of $\mathrm{p} = [u, v, 1]^T$ can be obtained after light strips detection. From Equation (1), the direction vector of Cp in {C} is:

$$\mathrm{s}^c = \mathrm{K}^{-1}\mathrm{p} = [m, n, 1]^T \tag{3}$$

where $m = \frac{(u-u_c)dx}{f}$, $n = \frac{(v-v_c)dy}{f}$.

Then the coordinates of P in {C} can be represented as:

$$\mathrm{P}^c = \begin{bmatrix} \lambda \mathrm{s}^c \\ 1 \end{bmatrix} \tag{4}$$

where λ are variables that need to be calculated next.

In the cross-structured light frame, assume that point P is on the vertical structured light plane, which means that $x^l = 0$; Equation (2) can be represented as follows [17]:

$$\mathrm{P}^c = \begin{bmatrix} {}^C_L r_{12} y^l + {}^C_L r_{13} z^l + {}^C_L t_1 \\ {}^C_L r_{22} y^l + {}^C_L r_{23} z^l + {}^C_L t_2 \\ {}^C_L r_{32} y^l + {}^C_L r_{33} z^l + {}^C_L t_3 \\ 1 \end{bmatrix} \tag{5}$$

Equation (6) can be obtained by combining Equations (4) and (5):

$$\mathrm{A} \begin{bmatrix} y^c \\ z^c \\ \lambda \end{bmatrix} = -{}^C_L \mathrm{T} \tag{6}$$

where $\mathrm{A} = \begin{bmatrix} {}^C_L r_{12} & {}^C_L r_{13} & -m \\ {}^C_L r_{22} & {}^C_L r_{23} & -n \\ {}^C_L r_{32} & {}^C_L r_{33} & -1 \end{bmatrix}$.

Then, λ can be calculated by solving Equation (6):

$$\lambda = \frac{A {}^C_L t_1 + B {}^C_L t_2 + C {}^C_L t_3}{\det(\mathrm{A})} \tag{7}$$

where $A = {}^C_L r_{22} {}^C_L r_{33} - {}^C_L r_{32} {}^C_L r_{23}$, $B = {}^C_L r_{13} {}^C_L r_{32} - {}^C_L r_{12} {}^C_L r_{33}$ and $C = {}^C_L r_{12} {}^C_L r_{23} - {}^C_L r_{13} {}^C_L r_{22}$.

Substituting (7) into (4), the coordinates of any structured light points in the camera frame can be calculated. This is the same for coordinates of points on the horizontal structured light plane, which means that $y^l = 0$ can be calculated too.

Finally, the coordinates of all the structured light points in the camera frame are obtained by one measurement (as shown in Figure 4). The next key problem is to register two adjacent measurements while the system is scanning and to transform the points in the camera frame to the world frame. A highly accurate registration method based on multi-sensors is proposed in the next part.

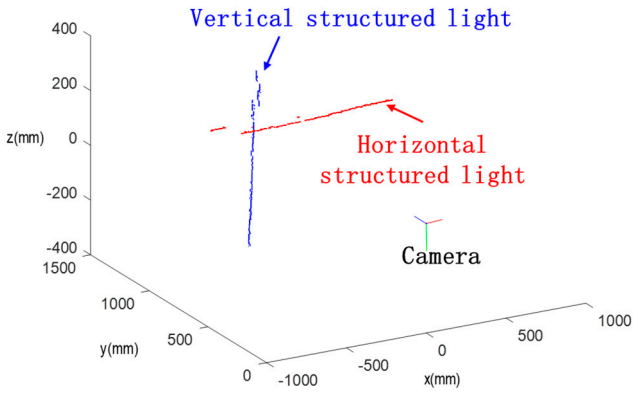


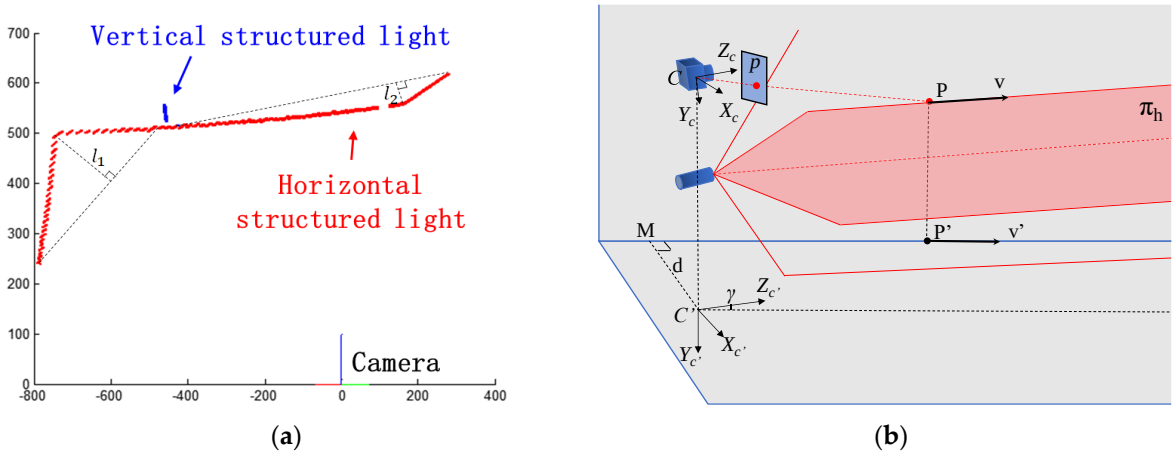
Figure 4. The coordinates of all the structured light points in the camera frame. The blue dots and red dots represent the data obtained from vertical structured light measurement and horizontal structured light measurement, respectively.

3.3. Heading Angle Measurement-Based Multi-Sensors

Once the coordinates of structured light points on two adjacent images are obtained, the registration of these two point clouds need to proceed. While a vertical structured light is used to scan the surroundings, a horizontal structure light is used to estimate the heading angle to the wall. In order to improve the robustness and accuracy of the heading angle estimation, the measurements from IMU are fused with an SLV sensor by a Kalman Filter.

3.3.1. Heading Angle Measurement Based on an SLV Sensor

Horizontal structured light is used to estimate the heading angle with respect to the wall. In order to detect numerous parts of the wall, distances l_1 and l_2 from the furthest measured point to the line created by the two endpoints and middle point, respectively, are estimated, as shown in Figure 5a.



(**a**) (**b**)

Figure 5. Heading angle measurement by using horizontal structured light. (**a**) Wall detection in vertical view. The blue dots and red dots represent the wall outline obtained from vertical structured light measurement and horizontal structured light measurement, respectively; (**b**) schematic of heading angle measurement.

If l_1 and l_2 are both bigger than a threshold, it means that three portions of the wall are detected. If only one distance is bigger than the threshold, it means two portions of the wall are detected. If none is bigger than the threshold, it means there is only one portion of the wall detected. For each found wall, the measured points on the wall are fitted to obtain the equation of the intersection line of the structured light plane and the wall.

Then, as Figure 5b shows, the intersection line in the camera frame can be represented by point P and vector v [29]:

$$P = \begin{bmatrix} x_P^c \\ y_P^c \\ z_P^c \end{bmatrix}, \quad v = \begin{bmatrix} x_v^c \\ y_v^c \\ z_v^c \end{bmatrix} \tag{8}$$

Consequently, the projection of the intersection line on the ground can be represented as:

$$P' = {}_c^{c'}\mathrm{R}P = \begin{bmatrix} x_P^{c'} \\ y_P^{c'} \\ z_P^{c'} \end{bmatrix}, \quad v' = {}_c^{c'}\mathrm{R}v = \begin{bmatrix} x_v^{c'} \\ y_v^{c'} \\ z_v^{c'} \end{bmatrix} \tag{9}$$

where ${}_c^{c'}\mathrm{R}$ is the rotation matrix from {C} to {C'}, respectively, which can be calculated through IMU because there is no change of heading angle from {C} to {C'} and the IMU can provide the precise value of pitch and roll.

In fact, due to the projection of the intersection line being on the ground, the projection of the intersection line can be represented by the equation of a 2D line:

$$z = ax + b \tag{10}$$

where $a = \frac{z_v^{c'}}{x_v^{c'}}$, and $b = z_P^{c'} - ax_P^{c'}$.

Then the heading angle with respect to the wall can be represented as:

$$\gamma = \arctan\left(\frac{z_v^{c'}}{x_v^{c'}}\right) \tag{11}$$

3.3.2. State Equations of the INS/SLV System

In order to estimate the heading angle change for higher accuracy, the measurements of INS are fused with the SLV sensor by a Kalman Filter, which can estimate the state of the system from a series of measurements observed over time, containing statistical noise and other inaccuracies [49]. Ideally, the true value of the heading angle change from moment k to k + 1 can be predicted using INS as follows [50]:

$$\Delta\gamma = (\omega_z - b_z)T \tag{12}$$

where ω_z is the true value of gyroscope measurements, b_z is the gyroscope bias and T is the sampling period. Although the gyroscope bias b_z can be obtained from the datasheet, b_z in the datasheet is a statistical value, which means there is an error in gyroscope bias when the INS is working. Thus, the heading angle change measured by INS can be written as:

$$\Delta\gamma^{INS} = (\omega_z - b_z - \delta b_z)T = \Delta\gamma - \delta b_z T \tag{13}$$

where δb_z is the error of gyroscope bias, which can be seen as an arbitrary constant. Then the error of heading angle change can be represented as:

$$\delta\Delta\gamma = \Delta\gamma^{INS} - \Delta\gamma = -\delta b_z T \tag{14}$$

By differentiating Equation (14) with respect to time, the following equation can be obtained:

$$\delta\Delta\dot{\gamma} = -\delta b_z \tag{15}$$

$$\delta \dot{b}_z = -\beta_z \delta b_z + \sqrt{2\beta_z \sigma_z^2} w \tag{16}$$

where β_z is the reciprocal of the correlation time constants of the random process associated with gyroscope measurements, σ_z^2 is the standard deviation of this random process and w is the zero mean Gaussian noise vector. The main purpose of the proposed system is to estimate $\delta\Delta\gamma$ and δb_z accurately, so the error state vector is defined as:

$$\delta x = \begin{bmatrix} \delta\Delta\gamma & \delta b_z \end{bmatrix}^T \tag{17}$$

The state equations of the INS/SLV system for the KF can be written as [50]:

$$\delta \dot{x} = \mathrm{F}\delta x + \mathrm{G}w \tag{18}$$

where δx is the error state vector, $\mathrm{F} = \begin{bmatrix} 0 & -1 \\ 0 & -\beta_z \end{bmatrix}$ is the transition matrix and $\mathrm{G} = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{2\beta_z\sigma_z^2} \end{bmatrix}$ is the noise matrix.

3.3.3. INS/SLV Measurement Model

Once the heading angle with respect to the wall is estimated by Equation (11), the heading angle change measured by the SLV sensor can be obtained:

$$\Delta\gamma^{SLV} = \gamma_k^{SLV} - \gamma_{k-1}^{SLV} = \Delta\gamma + \delta\gamma^{SLV} \tag{19}$$

where γ_k^{SLV} and γ_{k-1}^{SLV} are the heading angles measured by SLV at moments k and $k-1$, respectively, $\Delta\gamma$ is the true value of the heading angle change and $\delta\gamma^{SLV}$ is the SLV sensor measurement error which is assumed to be the zero mean Gaussian noise.

Then measurement z can be represented as:

$$z = \Delta\gamma^{SLV} - \Delta\gamma^{INS} = \mathrm{H}\delta x + v \tag{20}$$

where $\mathrm{H} = \begin{bmatrix} -1 & 0 \end{bmatrix}$ is the measurement matrix and $v = \delta\gamma^{SLV}$ is the measurement noise.

3.3.4. Process of Filtering

To apply the KF on the computer, the equations of the filter need to be discretized. The discrete state equations and measurement equations are [50]:

$$\begin{cases} \mathrm{X}_k = \Phi_{k,k-1}\mathrm{X}_{k-1} + \Gamma_{k-1}W_{k-1} \\ \mathrm{Z}_k = \mathrm{H}_k\mathrm{X}_k + V_k \end{cases} \tag{21}$$

where $\Phi_{k,k-1} = \mathrm{I} + \mathrm{F}(t_k)T$ is the discrete state transition matrix and $\Gamma_{k-1} = \mathrm{G}(t_{k-1})T$ is the discrete noise matrix.

Then the estimation of state vector $\hat{\mathrm{X}}_k$ can be solved as follows:

(1) Time update, which contains a state prediction (22) and mean square error prediction (23):

$$\hat{\mathrm{X}}_{k,k-1} = \Phi_{k,k-1}\hat{\mathrm{X}}_{k-1} \tag{22}$$

$$\mathrm{P}_{k,k-1} = \Phi_{k,k-1}\mathrm{P}_{k-1}\Phi_{k,k-1}^T + \Gamma_{k-1}\mathrm{Q}_{k-1}\Gamma_{k-1}^T \tag{23}$$

(2) Measurement update, which contains a state estimation (24), filter gain (25) and mean square error estimation (26):

$$\hat{\mathrm{X}}_k = \hat{\mathrm{X}}_{k,k-1} + \mathrm{K}_k(\mathrm{Z}_k - \mathrm{H}_k\hat{\mathrm{X}}_{k,k-1}) \tag{24}$$

$$\mathrm{K}_k = \mathrm{P}_{k,k-1}\mathrm{H}_k^T\left[\mathrm{H}_k\mathrm{P}_{k,k-1}\mathrm{H}_k^T + \mathrm{U}_k\right] \tag{25}$$

$$\mathrm{P}_k = [\mathrm{I} - \mathrm{K}_k\mathrm{H}_k]\mathrm{P}_{k,k-1}[\mathrm{I} - \mathrm{K}_k\mathrm{H}_k]^T + \mathrm{K}_k\mathrm{U}_k\mathrm{K}_k^T \tag{26}$$

where Q_{k-1} is the system noise covariance defined by system noise W_{k-1} and U_k is the measurement noise covariance defined by measurement noise V_k.

4. Results Analysis

The ranging accuracy of the proposed system is tested first. The factors that influence the ranging accuracy are analyzed and the ranging accuracies at different distances are tested. Then the indoor environments such as a closed room and corridors are reconstructed by the proposed system.

4.1. Experimental Setup

The environment reconstruction system is shown in Figure 6a. The GoPro Hero 4 Session, which is used as the monocular visual sensor, can provide a large field of view (red coordinate). The INS is manufactured by Advanced Navigation, and it can provide a three-axis angular rate, linear acceleration and earth magnetic field data (blue coordinate). The structured light projector can generate vertical and horizontal structured light, respectively (green coordinate). When the system rotates with an MAV, the vertical structured light can scan around the scene and the horizontal structured light will register the scan data between two measurements.



(**a**) (**b**) (**c**)

Figure 6. Experimental setup. (**a**) SLV/INS system. (**b**) Total station for the true ranging data. (**c**) Scan station for the true heading angle and environment reconstruction data.

The ranging accuracy analysis is carried out first. The true values of the distances are measured by a Leica Nova TS50 total station (shown in Figure 6b), which can measure distances with an accuracy of about 0.1 mm at a distance of hundreds of meters.

Then the heading angle measurement and the environment reconstruction accuracy are analyzed. To evaluate the performance of the heading angle measurement and reconstruction, the heading angle and points cloud data using a Leica P16 scan station (maximum scanning range is 80 m; distance ranging accuracy is 0.5 mm; and angle measurement accuracy is 1.5″, shown in Figure 6c) are compared as a benchmark.

4.2. Ranging Accuracy Analysis

In terms of Equations (4) to (7), the ranging accuracy is decided by the rotation matrix and translation matrix between the structured light frame and the camera frame, as well as the coordinates of light strips in the image frame. In the proposed system, the structured light frame is parallel to the camera frame. Therefore, the influence of the translation matrix and image resolution will be analyzed.

Distances to a wall, from 500 mm to 4000 mm with intervals of 500 mm, are measured by the SLV system. Distances from the camera center to the wall are measured by all the

structured light points on the images. The true values of the distances are measured by a total station. Then the mean error and the Root Mean Square Error (RMSE) are calculated. The mean error of the ranging mainly reflects the deviation of the SLV sensor's calibration and light strips detection, and the RMSE of the ranging reflects the random error of the light strips detection.

4.2.1. Influence of the Translation Matrix

Three cases of the proposed system are set out to evaluate the influence of the translation matrix on the distance measurement. All of these three cases have same image resolution. The distance from the camera center to the vertical structured light plane in Case A2 is nearly double that in Case A1 and is nearly half of that in Case A3. The distance from the camera center to the horizontal structured light plane and rotation matrix in each case remain unchanged. In each case, the SLV sensor is calibrated after setting because there may be tiny variations in the rotation matrix after each setting. Table 1 shows the parameters of the three cases.

Table 1. Three cases of different translation matrices.

	Image Resolution	Translation Matrix (mm)	Rotation Matrix
Case A1	1920 × 1080	$\begin{bmatrix}113.3 & 78.6 & 0.0\end{bmatrix}$	$\begin{bmatrix}1.000 & -0.009 & 0.000\\ 0.003 & 1.000 & -0.001\\ 0.000 & 0.002 & 1.000\end{bmatrix}$
Case A2	1920 × 1080	$\begin{bmatrix}247.9 & 78.9 & 0.0\end{bmatrix}$	$\begin{bmatrix}1.000 & -0.004 & 0.000\\ 0.006 & 1.000 & -0.001\\ 0.000 & 0.001 & 1.000\end{bmatrix}$
Case A3	1920 × 1080	$\begin{bmatrix}425.6 & 78.8 & 0.0\end{bmatrix}$	$\begin{bmatrix}1.000 & -0.007 & 0.000\\ 0.002 & 1.000 & 0.000\\ 0.000 & 0.001 & 1.000\end{bmatrix}$

For each case, distances from 500 mm to 4000 mm with 500 mm intervals are measured by the SLV sensor. The mean errors and RMSEs of all the measurements in each case are shown in Figure 7.

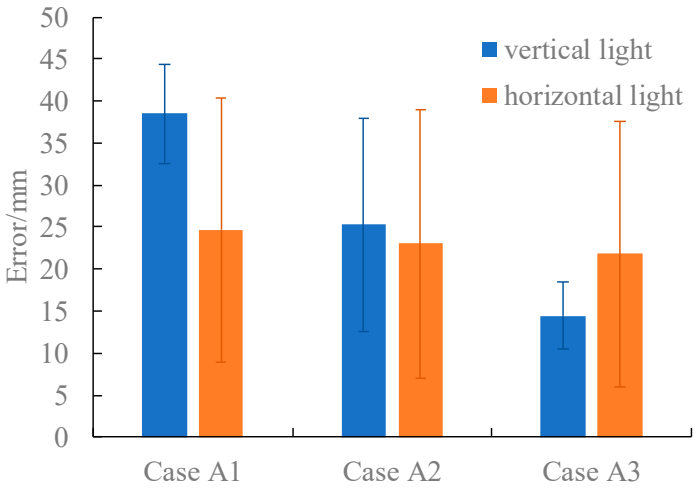


Figure 7. Comparison of mean errors and RMSEs with different translation matrices.

In Figure 7, the mean errors and RMSEs of horizontal light are almost the same because the distances between the camera center and the horizontal light plane are unchanged. This means that the accuracy of calibrations in the three cases are same. Then, the variation of error and RMSE of vertical light represents the fact that the accuracy of ranging is inversely proportional to the distance between the camera center and the structured light plane. However, there are also drawbacks of the larger distance between the camera center and

the structured light plane. When the distance between the camera center and structured light plane is large, the camera may not capture the light strips due to the limited field of view and the size of the SLV system is unsuitable for the MAV.

4.2.2. Influence of Image Resolution

Although the gray weight centroid method is used to detect the sub-pixel coordinates in the image frame, the accuracy of light strips detection is affected by the image resolution. Three cases with different resolution are set out to analyze the effect of image resolution, as shown in Table 2.

Table 2. Three cases of different image resolutions.

	Image Resolution	Translation Matrix (mm)	Rotation Matrix
Case B1	640 × 480	$\begin{bmatrix}247.9 & 78.9 & 0.0\end{bmatrix}$	$\begin{bmatrix}1.000 & -0.004 & 0.000\\ 0.006 & 1.000 & -0.001\\ 0.000 & 0.001 & 1.000\end{bmatrix}$
Case B2	1920 × 1080	$\begin{bmatrix}247.9 & 78.9 & 0.0\end{bmatrix}$	$\begin{bmatrix}1.000 & -0.004 & 0.000\\ 0.006 & 1.000 & -0.001\\ 0.000 & 0.001 & 1.000\end{bmatrix}$
Case B3	6000 × 4000	$\begin{bmatrix}247.9 & 78.9 & 0.0\end{bmatrix}$	$\begin{bmatrix}1.000 & -0.004 & 0.000\\ 0.006 & 1.000 & -0.001\\ 0.000 & 0.001 & 1.000\end{bmatrix}$

For each case, distances from 500 mm to 4000 mm with 500 mm intervals are measured by the SLV system. The mean errors and RMSEs of all the measurements in each case are shown in Figure 8.

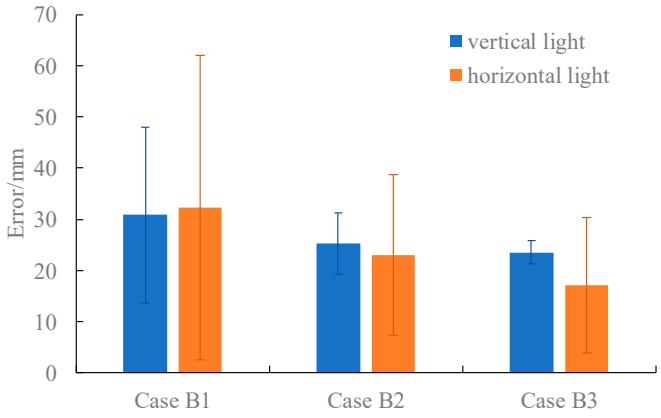


Figure 8. Comparison of mean errors and RMSEs with different image resolution.

In Figure 8, it can be seen that the higher the image resolution is, the lower the mean error and RMSE are. The RMSE, which is caused by the random error of strips detection, is influenced enormously by the resolution. The mean error, which is caused by the fixed deviation of strips detection, is influenced less. This means that the random error in strips detection accounts for the main component. Improving the resolution is the way to reduce the error of strips detection. However, a higher resolution of an image costs more time to calculate.

4.2.3. Influence of Measured Distance

Through the above analysis, the parameters of the SLV system in Table 3 are chosen for the environment reconstruction. In fact, the choice of the parameters is flexible for better accuracy or for faster calculation. Before the environment reconstruction, the relationship between the ranging accuracy and measured distance is analyzed in this part.

Table 3. The parameters of SLV sensor.

Image Resolution	Translation Matrix (mm)	Rotation Matrix
1920 × 1080	$[247.9 \quad 78.9 \quad 0.0]$	$\begin{bmatrix} 1.000 & -0.004 & 0.000 \\ 0.006 & 1.000 & -0.001 \\ 0.000 & 0.001 & 1.000 \end{bmatrix}$

A comparison of the mean error and RMSE at different distances is shown in Figure 9.

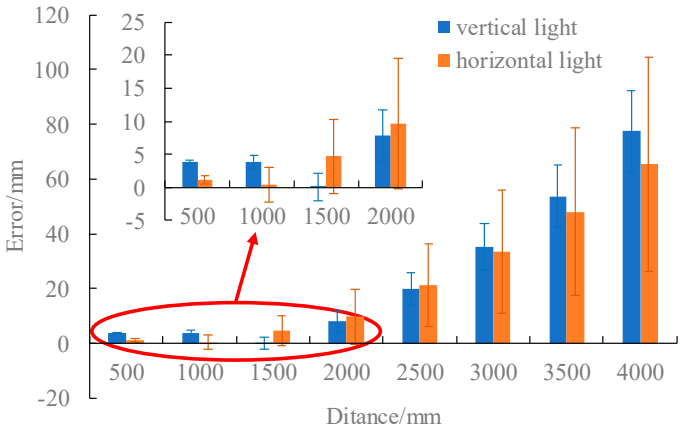


Figure 9. Comparison of the mean error and RMSE at different distances.

In Figure 9, when the distances are less than 2 m, the mean errors of ranging are less than 10 mm for both vertical and horizontal structured light. When the distances are greater than 2.5 m, the mean errors increase rapidly. This is because calibration of the SLV sensor is proceeded with at a distance of 1 m to 1.5 m, so the structured light planes are fitted more accurately at these distances. The accuracy can be improved when calibration is proceeded with at longer distances, but the corners on the chessboard will be hard to detect unless a much bigger chessboard is used. In fact, as shown in Table 4, the percentages of ranging accuracy are all less than 2% and this is accurate enough for environment reconstruction of indoor navigation.

Table 4. Percentage of ranging accuracy at different distances.

Distance/mm	500	1000	1500	2000	2500	3000	3500	4000
Accuracy of vertical light	0.78%	0.39%	0.01%	0.40%	0.79%	1.17%	1.54%	1.94%
Accuracy of horizontal light	0.23%	0.04%	0.32%	0.49%	0.85%	1.12%	1.37%	1.64%

On the other hand, the RMSEs of ranging for two structured light planes both increase with increased distance in Figure 8. When the distances are less than 2.5 m, the RMSEs of ranging are less than 4 mm for vertical structured light and are less than 10 mm for horizontal structured light. The reason for is that the distance between the camera center and the horizontal structured light plane is less than that of the vertical structured light plane, so that the same pixel error will cause more error in ranging for horizontal structured light.

4.3. Environment Reconstruction Analysis

Measurements of the 3D laser scan station are used as ground truth to test the performance of the proposed environment reconstruction system. A closed environment such as a kitchen and an open environment such as a corridor are reconstructed in the experiments. To evaluate the performance of reconstruction, the points clouds data using a Leica scan station are compared as a benchmark. The software *Cyclone* is used to obtain the points clouds from the scan station. The software *Cloud Compare* is used to estimate the accuracy of

the reconstruction results. For a clear display, all the ceilings and floors of the reconstruction results are removed.

The reconstruction result of a closed room is shown in Figure 10. A partial real environment is shown in Figure 10a. Constructions such as walls, corners and cabinets are clearly reconstructed in the points cloud. Details such as a refrigerator, microwave oven and toaster are also reconstructed in the points cloud. Figure 10b shows the heading angle changes between every two images. The curve "before filtering" means the heading angle change is measured by SLV only. It can be seen that the error is large, especially when the walls are separated into several parts. After fusing with INS, the outputs of the filter can register each measurement accurately. Figure 10c shows the distances between the points measured by SLV and the points measured by the 3D laser scan station. Figure 10d shows the error distribution of all the points. The horizontal axis represents the distance error and the vertical axis represents the number of points. It can be seen that about 94% of the points have an error of less than 50 mm. Only the window on the door and the edge of the sink have an error larger than 75 mm (the orange and red points in Figure 10c).

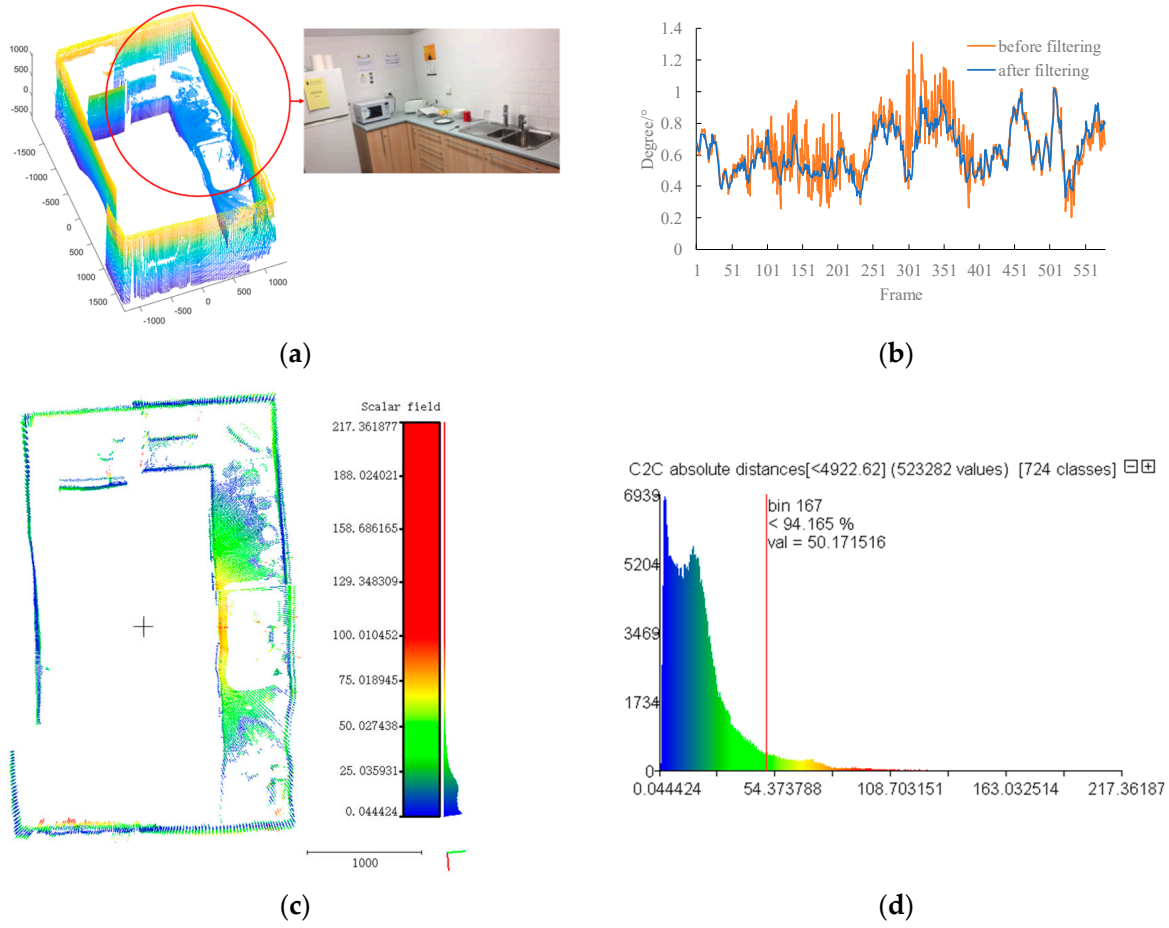


Figure 10. Reconstruction result of the kitchen. (**a**) Points cloud and partial real environment of the kitchen; (**b**) heading angle changes of the adjacent images for the kitchen; (**c**) top view of the kitchen; (**d**) statistical result of the kitchen.

For the corridor, it is hard to reconstruct the whole environment by one scan, due to the obstruction of the wall. The reconstruction is divided into two steps: (1) scanning the corridor at different locations respectively and (2) registering the points clouds from different locations. The reconstruction results are shown in Figure 11.

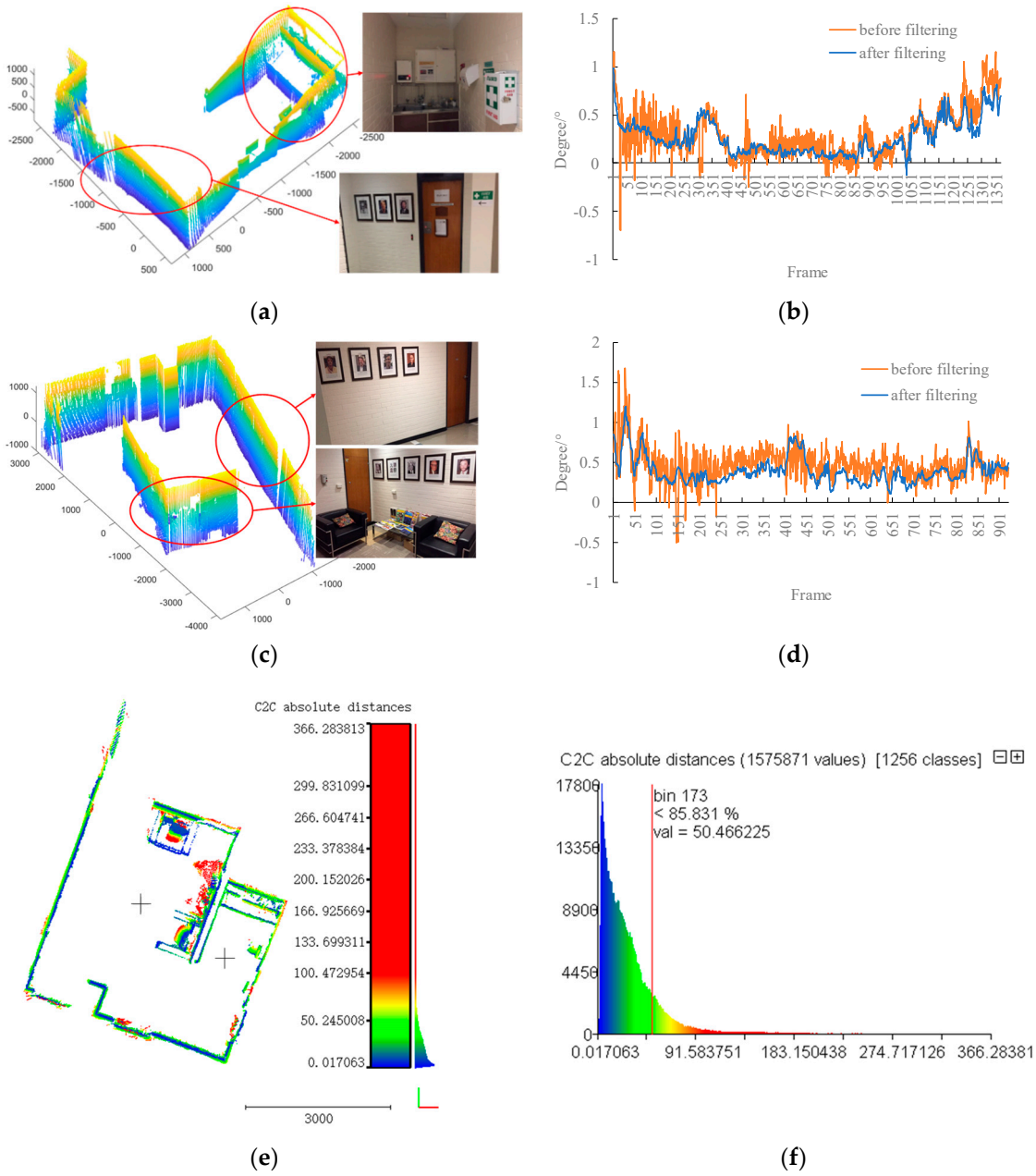


Figure 11. Reconstruction result of corridor. (**a**) Reconstruction result and partial real environment of corridor1; (**b**) heading angle changes of the adjacent images for corridor1; (**c**) error of each point in reconstructed environment of corridor2; (**d**) statistical result of reconstructed environment for corridor2; (**e**) top view of whole corridor; (**f**) statistical result of whole corridor.

Figure 11a,c shows the partial real environments and the results of the reconstruction. It can be seen that not only large objects such as the water heater and first-aid box are well reconstructed, but the photos on the wall can also be clearly seen. Figure 11b,d shows an estimation of the heading angle changes before and after filtering. The accuracy of angle estimation is improved by fusing the measurement of INS. In Figure 11e, the scanning results in different locations are registered by improved Iterative Closest Points (ICPs) [51]. The two black crosses represent the location of two scans. Although the glass table, the bolster on the sofa and the window on the door have an error larger than 100 mm, most of the points have an error of less than 50 mm. Figure 11f shows that about 85% of the points have an error of less than 50 mm.

Table 5 shows the mean error and standard deviation of two kinds of environment. The mean error and standard deviation of the kitchen are both less than 20 mm, and the mean error and standard deviation of the corridor are less than 30 mm. This is because the range of the corridor is larger and the glass objects such as the window and table are larger. Even so, the accuracy of both reconstruction results are accurate enough for MAV indoor navigation.

Table 5. Mean error and standard deviation of two kinds of environments.

	Kitchen	**Corridor**
Mean error (mm)	19.06	28.28
Standard deviation (mm)	16.92	29.52

5. Conclusions

Taking into consideration that the laser scan station is expensive and heavy for indoor environment reconstruction, a low-cost environment reconstruction system which consists of an SLV sensor and INS is proposed in this paper. The laser triangulation method is used for the SLV sensor to calculate coordinates of the points on structured light planes. The vertical structured light plane is used to scan the surroundings, and the horizontal structured light plane is used to estimate the angle change between two adjacent measurements. To improve the robustness and accuracy of the angle measurement, a KF with simple construction is designed to fuse the inertial information from INS.

To evaluate the performance of the proposed system, the ranging accuracy of the SLV sensor is analyzed first. The influences of the translation matrix and image resolution are tested to choose the most suitable parameter for the indoor environment reconstruction. The range of results for different distances shows that the ranging error is less than 10 mm while the distance is less than 2 m for both vertical and horizontal structured light, and both of them have an accuracy of less than 2% when the distance is less than 4 m. Then, an experiment of environment reconstruction is proceeded with. The reconstruction results of two kinds of environment show that both the main construction and some details can be reflected clearly, and the KF for the heading angle estimation is effective. Comparisons with the points cloud data from the scan station show that the error of most points is less than 50 mm for both kitchen (94%) and corridor (85%), and the mean error and standard deviation of kitchen and corridor are less than 20 mm and 30 mm, respectively. All the experimental results show that the proposed low-cost reconstruction system can provide indoor environments which are accurate enough for the navigation of unmanned systems.

The method proposed in this paper preliminarily demonstrates the feasibility and effectiveness of the low-cost 3D environment reconstruction system under indoor scenarios with low illumination and sparse features, relying on the INS of micro unmanned systems for 3D reconstruction. In further applications, it is necessary to introduce advanced AI technologies to optimize the robustness of the system to cope with low-structured indoor environments. Additionally, research on unmanned system control strategies applicable to this system is required to better support multiple point scans of the system in complex environments.

Author Contributions: Conceptualization and methodology, Y.W. and Q.M.; software, writing—original draft preparation and validation, Y.W.; investigation and data curation, B.D. and H.W.; visualization, Y.Z. and H.J.; writing—review and editing, supervision, project administration and funding acquisition, Q.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work is funded by the National Natural Science Foundation of China (Grant No. 62203111), Aeronautical Science Foundation of China (20220008069003), Natural Science Foundation of Jiangsu Province (No. BK20231434) and the Open Fund of State Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University (Grant No. 21P01).

Data Availability Statement: The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Acknowledgments: The authors appreciate all the valuable suggestions and discussions from Jinling Wang, University of New South Wales.

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

References

1. Jiang, S.; Liu, J.; Li, Y.; Weng, D.; Chen, W. Reliable Feature Matching for Spherical Images via Local Geometric Rectification and Learned Descriptor. *Remote Sens.* **2023**, *15*, 4954. [CrossRef]
2. Xu, Y.; Wan, D.; Bi, S.H.; Guo, H.; Zhuang, Y. A FIR filter assisted with the predictive model and ELM integrated for UWB based quadrotor aircraft localization. *Satell. Navig.* **2022**, *4*, 2. [CrossRef]
3. El-Sheimy, N.; Li, Y. Indoor navigation: State of the art and future trends. *Satell. Navig.* **2021**, *2*, 7. [CrossRef]
4. Kang, Z.; Yang, J.; Yang, Z.; Cheng, S. A Review of Techniques for 3D Reconstruction of Indoor Environments. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 330. [CrossRef]
5. Lu, Q.; Pan, Y.; Hu, L.; He, J. A Method for Reconstructing Background from RGB-D SLAM in Indoor Dynamic Environments. *Sensors* **2023**, *23*, 3529. [CrossRef] [PubMed]
6. Tran, H.; Khoshelham, K.; Kealy, A. Geometric comparison and quality evaluation of 3d models of indoor environments. *ISPRS J. Photogramm. Remote Sens.* **2019**, *149*, 29–39. [CrossRef]
7. Chiang, K.W.; Chiu, Y.T.; Srinara, S.; Tsai, M. Performance of LiDAR-SLAM-based PNT with initial poses based on NDT scan matching algorithm. *Satell. Navig.* **2023**, *4*, 3. [CrossRef]
8. Hasselbach, J.; Bogatscher, S.; Rembe, C. Laser scanner module with large sending aperture and inherent high angular position accuracy for three-dimensional light detecting and ranging. *Opt. Eng.* **2019**, *58*, 087101. [CrossRef]
9. Bosché, F.; Ahmed, M.; Turkan, Y.; Haas, C.T.; Haas, R. The value of integrating Scan-to-BIM and Scan-vs-BIM techniques for construction monitoring using laser scanning and BIM: The case of cylindrical MEP components. *Autom. Constr.* **2015**, *49*, 201–213. [CrossRef]
10. Andriasyan, M.; Moyano, J.; Nieto-Julián, J.E.; Antón, D. From Point Cloud Data to Building Information Modelling: An Automatic Parametric Workflow for Heritage. *Remote Sens.* **2020**, *12*, 1094. [CrossRef]
11. Ducke, B.; Score, D.; Reeves, J. Multiview 3D reconstruction of the archaeological site at Weymouth from image series. *Comput. Graph.* **2011**, *35*, 375–382. [CrossRef]
12. Kamble, T.U.; Mahajan, S.P. 3D Vision Using Multiple Structured Light-Based Kinect Depth Cameras. *Int. J. Image Graph.* **2022**, *24*, 2450001. [CrossRef]
13. Condotta, I.C.F.S.; Brown-Brandl, T.M.; Pitla, S.K.; Stinn, J.P.; Silva-Miranda, K.O. Evaluation of low-cost depth cameras for agricultural applications. *Comput. Electron. Agric.* **2020**, *173*, 105394. [CrossRef]
14. Fan, J.; Wang, X.; Zhou, C.; Ou, Y.; Jing, F.; Hou, Z. Development, calibration, and image processing of underwater structured light vision system: A survey. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 1–18. [CrossRef]
15. Han, Y.; Fan, J.; Yang, X. A structured light vision sensor for on-line weld bead measurement and weld quality inspection. *Int. J. Adv. Manuf. Technol.* **2020**, *106*, 2065–2078. [CrossRef]
16. Khan, D.; Cheng, Z.; Uchiyama, H.; Ali, S.; Asshad, M.; Kiyokawa, K. Recent advances in vision-based indoor navigation: A systematic literature review. *Comput. Graph.* **2022**, *104*, 24–45. [CrossRef]
17. Wang, Y.S.; Liu, J.Y.; Zeng, Q.H.; Liu, S. Visual pose measurement based on structured light for MAVs in non-cooperative environments. *Opt.-Int. J. Light Electron Opt.* **2015**, *126*, 5444–5451. [CrossRef]
18. Jamali, A.; Anton, F.; Rahman, A.A.; Boguslawski, P.; Gold, C.M. 3D indoor building environment reconstruction using calibration of rangefinder data. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2015**, *2*, 29–34. [CrossRef]
19. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [CrossRef]
20. Rosten, E.; Porter, R.; Drummond, T. Faster and better: A machine learning approach to corner detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 105–119. [CrossRef] [PubMed]
21. Gui, S.; Song, S.; Qin, R.; Tang, Y. Remote Sensing Object Detection in the Deep Learning Era—A Review. *Remote Sens.* **2024**, *16*, 327. [CrossRef]
22. Snavely, N.; Seitz, S.M.; Szeliski, R. Photo tourism: Exploring photo collections in 3D. *ACM Trans. Graph.* **2006**, *25*, 835–846. [CrossRef]
23. Dewangan, K.; Saha, A.; Vaiapury, K.; Dasgupta, R. 3D Environment Reconstruction Using Mobile Robot Platform and Monocular Vision. In *Advanced Computing and Communication Technologies: Proceedings of the 9th ICACCT*; Springer: Singapore, 2016.
24. Phan, T.B.; Trinh, D.H.; Wolf, D.; Daul, C. Optical flow-based structure-from-motion for the reconstruction of epithelial surfaces. *Pattern Recognit.* **2020**, *105*, 107391. [CrossRef]
25. Bai, L.; Li, Y.; Kirubarajan, T.; Gao, X. Quadruple tripatch-wise modular architecture-based real-time structure from motion. *Neurocomputing* **2022**, *480*, 169–191. [CrossRef]
26. Zhang, C.; Du, S.; Xue, J.; Qi, X. Improved ICP algorithm with bounded rotation angle for 2d point set registration. In Proceedings of the First International Conference on Cognitive Systems and Information Processing, Beijing, China, 15–17 December 2012.

27. Wang, Y.; Zhou, T.; Li, H.; Tu, W.; Xi, J.; Liao, L. Laser point cloud registration method based on iterative closest point improved by Gaussian mixture model considering corner features. *Int. J. Remote Sens.* **2022**, *43*, 932–960. [CrossRef]
28. Vlaminck, M.; Luong, H.; Goeman, W.; Philips, W. 3D Scene Reconstruction Using Omnidirectional Vision and LiDAR: A Hybrid Approach. *Sensors* **2016**, *16*, 1923. [CrossRef]
29. Bosche, F. Automated recognition of 3d cad model objects in laser scans and calculation of as-built dimensions for dimensional compliance control in construction. *Adv. Eng. Inform.* **2010**, *24*, 107–118. [CrossRef]
30. Gutmann, E.D.; Larson, K.M.; Williams, M.W.; Nievinski, F.G.; Zavorotny, V. Snow measurement by GPS interferometric reflectometry: An evaluation at Niwot Ridge, Colorado. *Hydrol. Process.* **2012**, *26*, 2951–2961. [CrossRef]
31. Eitel, J.U.; Vierling, L.A.; Magney, T.S. A lightweight, low cost autonomously operating terrestrial laser scanner for quantifying and monitoring ecosystem structural dynamics. *Agric. For. Meteorol.* **2013**, *180*, 86–96. [CrossRef]
32. Feng, Y.; Wu, R.; Liu, X.; Chen, L. Three-Dimensional Reconstruction Based on Multiple Views of Structured Light Projectors and Point Cloud Registration Noise Removal for Fusion. *Sensors* **2023**, *23*, 8675. [CrossRef] [PubMed]
33. Kholodilin, I.; Li, Y.; Wang, Q. Omnidirectional vision system with laser illumination in a flexible configuration and its calibration by one single snapshot. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 9105–9118. [CrossRef]
34. Zhang, Z.; Zhou, H.; Wang, S.; Xu, C.; Lv, Y. Design and Research of Low-Cost and Self-Adaptive Terrestrial Laser Scanning for Indoor Measurement Based on Adaptive Indoor Measurement Scanning Strategy and Structural Characteristics Point Cloud Segmentation. *Adv. Civ. Eng.* **2022**, *2022*, 5681771. [CrossRef]
35. Gu, F.; Song, Z.; Zhao, Z. Single-Shot Structured Light Sensor for 3D Dense and Dynamic Reconstruction. *Sensors* **2020**, *20*, 1094. [CrossRef] [PubMed]
36. Mildenhall, B.; Srinivasan, P.; Tancik, M.; Barron, J.; Ramamoorthi, R.; Ng, R. Nerf: Representing scenes as neural radiance fields for view synthesis. *Commun. ACM* **2021**, *65*, 99–106. [CrossRef]
37. Barron, J.T.; Mildenhall, B.; Verbin, D.; Srinivasan, P.P.; Hedman, P. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 5470–5479.
38. Verbin, D.; Hedman, P.; Mildenhall, B.; Zickler, T.; Barron, J.T.; Srinivasan, P.P. Ref-nerf: Structured view-dependent appearance for neural radiance fields. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 5481–5490.
39. Tancik, M.; Weber, E.; Ng, E.; Li, R.; Yi, B.; Wang, T.; Kristoffersen, A.; Austin, J.; Salahi, K.; Ahuja, A.; et al. Nerfstudio: A modular framework for neural radiance field development. In *ACM SIGGRAPH 2023 Conference Proceedings*; ACM: New York, NY, USA, 2023; pp. 1–12.
40. Brossard, M.; Barrau, A.; Bonnabel, S. AI-IMU dead-reckoning. *IEEE Trans. Intell. Veh.* **2020**, *5*, 585–595. [CrossRef]
41. Liu, W.; Caruso, D.; Ilg, E.; Dong, J.; Mourikis, A.I.; Daniilidis, K.; Kumar, V.; Engel, J.; Valada, A.; Asfour, T. Tlio: Tight learned inertial odometry. *IEEE Robot. Autom. Lett.* **2020**, *5*, 5653–5660. [CrossRef]
42. Sodhi, P.; Kaess, M.; Mukadam, M.; Anderson, S. Learning tactile models for factor graph-based estimation. In Proceedings of the 2021 IEEE International Conference on Robotics and Automation (ICRA), Xi'an, China, 30 May–5 June 2021; pp. 13686–13692.
43. Baikovitz, A.; Sodhi, P.; Dille, M.; Kaess, M. Ground encoding: Learned factor graph-based models for localizing ground penetrating radar. In Proceedings of the 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Prague, Czech Republic, 27 September–1 October 2021; pp. 5476–5483.
44. Nam, D.V.; Gon-Woo, K. Learning Type-2 Fuzzy Logic for Factor Graph Based-Robust Pose Estimation with Multi-Sensor Fusion. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 3809–3821. [CrossRef]
45. Zhang, Z. A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 1330–1334. [CrossRef]
46. Bouguet. Camera Calibration Toolbox for Matlab. Available online: http://www.vision.caltech.edu/bouguetj/calib_doc/index.html#examples. (accessed on 1 August 2013).
47. Bi, D.X.; Liu, F.T.; Xue, Q.; Li, Z.G. New structured light vision sensor field calibration approach based on laser intersection lines. *Chin. J. Sci. Instrum.* **2009**, *30*, 1697–1701.
48. Demi, M. On the gray-level central and absolute central moments and the mass center of the gray-level variability in low-level image processing. *Comput. Vis. Image Underst.* **2005**, *97*, 180–208. [CrossRef]
49. Meng, Q.; Hsu, L.T. Resilient interactive sensor-independent-update fusion navigation method. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 16433–16447. [CrossRef]
50. Huang, G. Visual-inertial navigation: A concise review. In Proceedings of the 2019 International Conference on Robotics and Automation (ICRA), Montreal, QC, Canada, 20–24 May 2019; pp. 9572–9582.
51. Bergström, P.; Edlund, O. Robust registration of surfaces using a refined iterative closest point algorithm with a trust region approach. *Numer. Algorithms* **2017**, *74*, 755–779. [CrossRef]

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