sensors

MDPI

Article

Improving 3D Reconstruction Through RGB-D Sensor Noise Modeling

Fahira Afzal Maken [1,*,†], Sundaram Muthu [1,†], Chuong Nguyen [1], Changming Sun [1], Jinguang Tong [1,2], Shan Wang [1,2], Russell Tsuchida [1], David Howard [1], Simon Dunstall [1] and Lars Petersson [1]

[1] Data61, Commonwealth Scientific and Industrial Research Organisation (CSIRO), Canberra, ACT 2601, Australia; sundaram.muthu@data61.csiro.au (S.M.); chuong.nguyen@data61.csiro.au (C.N.); changming.sun@data61.csiro.au (C.S.); jinguang.tong@anu.edu.au (J.T.); shan.wang@anu.edu.au (S.W.); russell.tsuchida@data61.csiro.au (R.T.); david.howard@data61.csiro.au (D.H.); simon.dunstall@data61.csiro.au (S.D.); lars.petersson@data61.csiro.au (L.P.)

[2] School of Computing, Australian National University (ANU), Canberra, ACT 2601, Australia

* Correspondence: fahira.afzalmaken@data61.csiro.au

† These authors contributed equally to this work.

Abstract: High-resolution RGB-D sensors are widely used in computer vision, manufacturing, and robotics. The depth maps from these sensors have inherently high measurement uncertainty that includes both systematic and non-systematic noise. These noisy depth estimates degrade the quality of scans, resulting in less accurate 3D reconstruction, making them unsuitable for some high-precision applications. In this paper, we focus on quantifying the uncertainty in the depth maps of high-resolution RGB-D sensors for the purpose of improving 3D reconstruction accuracy. To this end, we estimate the noise model for a recent high-precision RGB-D structured light sensor called Zivid when mounted on a robot arm. Our proposed noise model takes into account the measurement distance and angle between the sensor and the measured surface. We additionally analyze the effect of background light, exposure time, and the number of captures on the quality of the depth maps obtained. Our noise model seamlessly integrates with well-known classical and modern neural rendering-based algorithms, from KinectFusion to Point-SLAM methods using bilinear interpolation as well as 3D analytical functions. We collect a high-resolution RGB-D dataset and apply our noise model to improve tracking and produce higher-resolution 3D models.

Keywords: sensor noise modeling; 3D reconstruction; RGB-D fusion

check for updates

Academic Editor: Miguel Oliveira

Received: 3 December 2024
Revised: 20 January 2025
Accepted: 21 January 2025
Published: 5 February 2025

Citation: Afzal Maken, F.; Muthu, S.; Nguyen, C.; Sun, C.; Tong, J.; Wang, S.; Tsuchida, R.; Howard, D.; Dunstall, S.; Petersson, L. Improving 3D Reconstruction Through RGB-D Sensor Noise Modeling. *Sensors* **2025**, *25*, 950. https://doi.org/10.3390/s25030950

Copyright: © 2025 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

Depth sensing is an essential component in perceiving the environment in three dimensions. With depth sensors providing rich 3D information, they have become integral to vision systems across manufacturing and other industries, facilitating various applications including 3D object segmentation, pose estimation [1], 3D reconstruction, and scene understanding [2]. Depth sensors enable the creation of accurate 3D models for printing, surface inspection, or part repair in reverse modeling. Specifically, advanced digital manufacturing applications including metrology, quality inspection, autonomous repair, 3D printing, and others require high-resolution scanning with low measurement uncertainty and submillimeter-level accuracy in the 3D reconstruction of the scanned objects. This is essential to capture fine details, achieve tight tolerances, meet dimensional specifications, or achieve the required surface finishing.

However, none of the raw depth values from the current depth sensors on the market are capable of achieving reliable submillimeter accuracy in the 3D reconstruction of scanned objects required for high-precision manufacturing applications. This limitation stems from noisy depth maps, characterized by high measurement uncertainty, which is the result of a combination of systematic errors (inherent to the system), and non-systematic errors (caused by uncontrolled environmental factors).

Systematic errors originate from the sensing methodology involved, such as structured light, time of flight, or stereo depth sensors [3]. Sensor calibration is typically performed to reduce the effect of systematic errors to some extent. Non-systematic errors originate from environmental influences such as the lighting conditions, the reflectivity of the target objects, measurement distance, and the angle between the object's surface normals and the principal axis of the camera. To this end, approaches in [3–8] have empirically modeled the noise characteristics of well-known low-cost RGB-D sensors such as Microsoft Kinect and Intel RealSense.

Zivid is a newer high-resolution 3D structured light camera whose noise characteristics have not yet been extensively studied. We focus on high-precision 3D scanning for manufacturing inspection applications, and this paper aims to model the noise characteristics of the Zivid sensor to optimize its utility in such applications. We quantify the noise in terms of axial noise (in the direction of the principal axis of the camera) and lateral noise (in the directions perpendicular to the axis of the camera), examining how these components vary with object depth and surface angle. Additionally, our study evaluates the influence of lighting conditions, exposure time, and the number of captures on the quality of depth maps obtained with the Zivid sensor (Zivid, Oslo, Norway).

In order to illustrate the use of our noise modeling to improve downstream tasks, we integrate our noise model with traditional methods such as KinectFusion and with modern neural-based approaches like Point-SLAM [9]. Filtering noisy depth values using our noise model not only improves the quality of 3D reconstruction but also captures high-resolution details, as demonstrated in Figure 1.

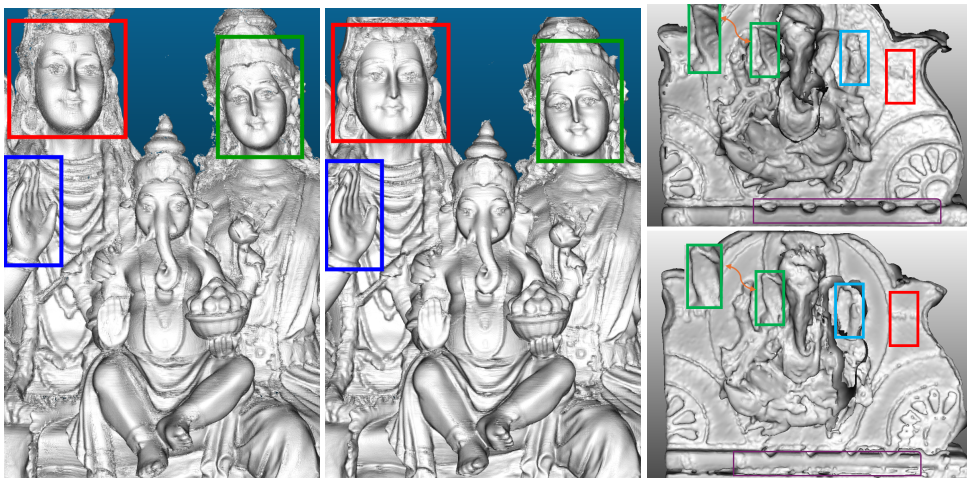


Figure 1. Our noise model, when integrated with KinectFusion [10], improves the quality of the surface estimation, highlighted in colored boxes (**middle**) compared with the quality without noise filtering (**left**). On the right, the noisy depth filtering using our noise model (**top right**) effectively captures high-resolution details, such as the ridges in the ear (green box) and the background texture (red box), which are absent in the unfiltered version (**bottom right**).

There is a notable absence of high-precision benchmark RGB-D datasets for submillimeter-level 3D reconstruction. To fill this gap, we introduce a new dataset captured with the Zivid camera mounted on a robotic arm. This dataset features objects with complex geometry,

scanned from multiple viewpoints achieved by rotating both the object on a turntable and the camera on a robot arm with high precision.

The main contributions of this paper are as follows:

- To improve 3D scanning in advanced manufacturing, we empirically model the axial and lateral noise characteristics of a high-resolution RGB-D Zivid sensor as a function of the measurement distance and surface angle of the scanned objects. We also provide insights on performance under different lighting conditions, exposure time, and different capture settings.
- We demonstrate how to employ our noise models in 3D reconstruction pipelines from traditional to neural-based methods to improve both the quality of 3D reconstruction and pose estimation using bilinear interpolation as well through fitted analytical functions.
- We collect and publish a new dataset that contains high-resolution RGB-D scans of objects with complex geometry that can be used in many applications including pose estimation and 3D reconstruction.

2. Related Work

In the following, we briefly review various noise characterization methods and their applications to improve computer vision tasks. We then provide an overview of the 3D reconstruction methods using RGB-D sensor data.

2.1. Noise Modeling of RGB-D Sensors

Systematic and non-systematic errors generally influence the accuracy of depth measurements from RGB-D sensors. The first detailed study of the noise characteristics of the well-known Kinect V1 RGB-D sensors was performed in [8]. In particular, the authors performed a calibration step to model the systematic errors and derived a theoretical random noise model to account for non-systematic errors. Nguyen et al. [5] present an empirical model for measuring the lateral and axial noise distributions of the Kinect V1 sensor related to the distance and angle from the surface. Similarly to Nguyen et al. [5], Choo et al. [11] characterize the noise model of the Kinect sensor in both the axial and lateral directions. In addition, Fankhauser et al. [6] characterize the influence of ambient light for indoor, overcast, and direct sunlight situations, considering the angle of incidence of sunlight. Tölgyessy et al. [4] extend the analysis to include the Azure Kinect, examining its noise characteristics during warming up, reflectivity levels, pixel positions, and in outdoor lighting conditions.

With the growing variety of RGB-D sensors in the market, several studies [3,7,12] compare the noise characteristics of different RGB-D sensors, including various versions of the Kinect. Mallick et al. [7] provide a comprehensive review of the methods that characterize the Kinect sensor noise. Additionally, they introduce and measure interference noise when multiple Kinect sensors are used. Jing et al. [3] compare the depth image quality and noise characteristics of the Kinect V1, Kinect V2, and the PrimeSense sensors. Halmetschlager et al. [12] compare the noise characteristics of ten different RGB-D sensors. They suggest employing structured light cameras for robots operating in uncontrolled conditions due to the comparatively lower noise variation compared with other sensor technologies.

While the previously mentioned methods rely on a calibration target object for noise characterization, some approaches [13–15] estimate noise or depth uncertainty directly from the captured data. Dryanovski et al. [15] propose a mixture-of-Gaussians convolution method to quantify the uncertainty in a single depth frame. Compared with the approaches in [5,8], Lu and Song [15] represent the uncertainty around the edges of the object more

accurately. Proenca et al. [14] propose a depth filter that combines approaches in [8,15] to model the uncertainty of spatial and temporal depth measurement. Dong et al. [13] propose a regularization technique that models depth uncertainty in the incremental Bayesian update step to better handle outliers.

In contrast to these preceding contributions, our focus is on modeling the noise characteristics of the high-resolution Zivid camera. The Zivid camera is well-suited for advanced manufacturing applications that require more precise measurements, but empirical noise studies for this particular sensor are lacking in the existing literature.

2.2. Application of Noise Models to Improve Computer Vision Tasks

Noisy depth maps degrade the quality of the scans and reduce the accuracy of downstream applications such as pose estimation, point cloud alignment, SLAM, robot perception, and 3D reconstruction. Hence, various studies have utilized sensor noise models to filter or fuse depth maps to enhance the accuracy of these tasks [5,8,16–25].

For SLAM applications, Gutierrez et al. [16,17] improve the dense visual odometry method (DVO SLAM) [26,27] by considering depth errors. Specifically, Gutierrez-Gomez et al. [16] parameterize geometric error using inverse depth, while dense noise aware SLAM (DNA-SLAM) [17] considers noise explicitly and uses a sophisticated weighting mechanism to improve odometry. Yamaguchi et al. [18] integrate the noise characteristics of the Kinect sensor into the iterative closest point (ICP) algorithm, improving the accuracy of SLAM. Furthermore, feature-based SLAM methods [21] use depth uncertainty models derived from sensors such as structured light cameras to improve odometry despite their lower depth resolution. Bingham procrustean alignment [24] uses uncertainty estimates from noisy depth images to extract reliable surface features for robust pose alignment.

A robot perception system in [20] employs the sensor noise model [5] to achieve optimal mapping quality. Dhawale et al. [28] employ the sensor noise model [14] to accurately compute the correspondences between the existing map and the sensor observation during incremental probabilistic map updates. In addition, incorporating sensor noise characteristics reduces the memory requirement of the map representation. Voxblox [29] improves mapping and path planning in unknown environments for autonomous flight applications by using truncated signed distance fields based on the noise model (TSDF) [5]. Saulnier et al. [25] decrease the uncertainty of the map in the autonomous exploration task by taking into account the noise based on the measured depth.

In relation to 3D reconstruction and other applications, Nguyen et al. [5] demonstrate that exploiting the noise model in the KinectFusion system [10] improves pose estimation accuracy and allows reconstruction of finer details, smaller objects, and thinner surfaces. Iversen et al. [23] exploit a noise model [5] for simulating Kinect V1 depth images, which can be used in various computer vision algorithms requiring a substantial dataset. Recently, Ran et al. [30] have introduced an implicit neural representation for path planning and active reconstruction, using the sensor noise model to assess reconstruction uncertainty to determine the next best view for active scanning. Katar et al. [31] model lateral and axial noise in 3D camera data to enhance synthetic training datasets, demonstrating that this improves neural network performance in real-world object segmentation tasks. Similarly, Cai et al. [32] quantify and model non-systematic noise in the PMD Flexx2 Time-of-Flight depth camera, proposing Gaussian-based models for axial and lateral noise to aid in accurate sensor simulation for robotic applications. In Rustler et al. [33], the authors evaluate the performance of four stereoscopic RGB-D cameras—Intel RealSense D435, Intel RealSense D455, StereoLabs ZED 2, and Luxonis OAK-D Pro—across various scenarios, providing insights into their suitability for different robotic applications. The research in [34] assesses the performance of three depth sensors—Intel RealSense D415, Intel

RealSense D405, and Stereolabs ZED-Mini—within close-range surgical simulation settings, highlighting their accuracy, fill rate, and suitability for reconstructing anatomical structures. Additionally, Keetha et al. [35] propose methods to enhance depth value approximation in 3D image-based modeling for complex objects by introducing noise filtering and inverse perspective mapping techniques, resulting in improved accuracy and processing speed.

2.3. 3D Reconstruction from Multi-View RGB-D Scans

Recent research in 3D reconstruction and scene understanding is broadly categorized into two distinct approaches. Initially, fusion-based methods were pioneered, drawing heavily on the foundational KinectFusion algorithm introduced by Izadi et al. [10]. This line of work was significantly advanced by Nießner et al. [36], who facilitated real-time, large-scale reconstruction through voxel hashing. Subsequently, DynamicFusion [37] expanded these capabilities to dynamic scenes via a novel warp field technique. MaskFusion [38] introduced an innovative approach to segment, track, and reconstruct moving objects by integrating semantic segmentation. BundleFusion [39] further refined these methodologies by proposing a comprehensive system that ensures real-time, globally consistent 3D reconstructions through on-the-fly optimization and surface re-integration. On the other hand, the exploration of neural radiance field (NeRF)-based methods represents a shift towards leveraging deep learning for scene reconstruction. GO-SLAM [40] employs an SDF-based NeRF for scene reconstruction, whereas Point-SLAM [9] utilizes Point-NeRF [41] for similar purposes. Our noise model improves 3D reconstruction accuracy in both fusion-based and neural network methods.

3. Noise Modeling

3.1. Noise Characteristics

Depth measurement errors in structured light sensors can occur due to sensor warm-up, and varying depth distances, angles, binning modes, reflectivity, and distortion. In this paper, we characterize the noise in three dimensions as a function of depth distance, z, and angles, θ, similar to [5]. Specifically, we quantify the noise characteristics in terms of axial noise, $\sigma_Z = f(z, \theta) : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$, which denotes noise along the principal axis of the camera, and lateral noise, $\sigma_L = g(z, \theta) : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$, representing noise in directions perpendicular to the camera axis, as shown in Figure 2.

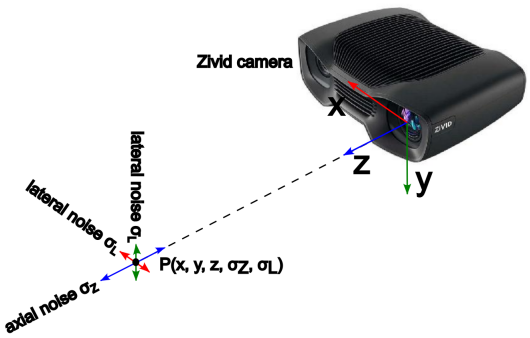


Figure 2. Illustration of Zivid camera noise components at an arbitrary point, $P(x, y, z)$, measured by the camera. Axial noise, σ_Z, and lateral noise, σ_L, represent the uncertainty of the measured location of point P, along with the z-axis and x- and y-axes, respectively.

3.2. Experimental Setup

Figure 3a shows the experimental setup used to collect data to estimate the noise functions f and g. It consists of a robot arm, camera, and a planar target. We use the Zivid 2 (https://www.zivid.com/zivid-2, accessed on 20 January 2025) structured light RGB-D camera whose noise model is to be estimated, as shown in Figure 3d. We use a

planar target (a calibration board (https://calib.io/, accessed on 20 January 2025) that has a dimensional accuracy of 10 µm) that is mounted on a tripod, as shown in Figure 3c. The planar target and the axis of rotation of the planar target are both perpendicular to the principal z-axis of the camera. Additionally, the camera is aligned to focus on the center of the plane. The target is capable of rotating freely around the vertical axis. The camera is mounted on the UR5e robot arm (https://www.universal-robots.com/products/ur5-robot/, accessed on 20 January 2025) to automate data collection precisely. An eye-in-hand calibration step is performed to determine the transformation between the camera and the robot base, ensuring accurate pose estimates of the Zivid camera within a global coordinate system. We vary the following parameters in our setup to collect data to model various noise characteristics of the sensor.

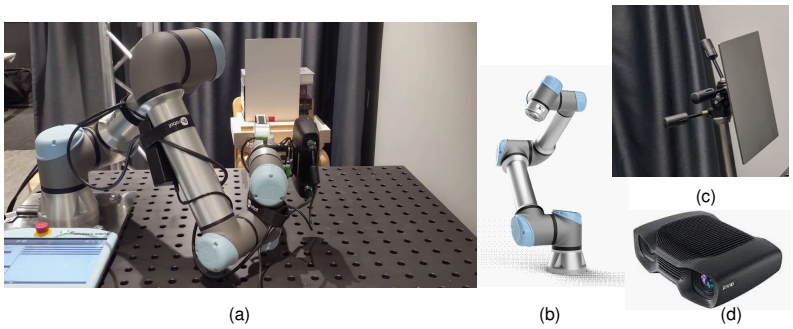


Figure 3. Our experimental setup for modeling sensor noise includes a robot arm (**b**), a planar target (**c**), and a Zivid 2 RGB-D structured light sensor (**d**), as shown in (**a**).

- *Distance z* : the distance between the target and the camera, z, is varied from 37.5 cm to 107.0 cm in steps of 2.5 cm. This axial movement of the camera is performed by moving the robot arm towards and away from the target in a straight line perpendicular to the planar target.
- *Angle θ*: the angle of the planar target is varied from 0° to 90° in steps of 10° with rotation perpendicular to the camera's principal axis.
- *Lighting conditions*: the data collection process involves scans of the calibration board under both light and dark conditions at 0° for all distances.
- *Exposure time*: the exposure time or the duration of the shutter opening time is varied from 1677 ms to 100 s.
- *Number of captures*: the number of captures is varied from 1 to 5, combining multiple acquisitions with different apertures to enable capture at a high dynamic range.

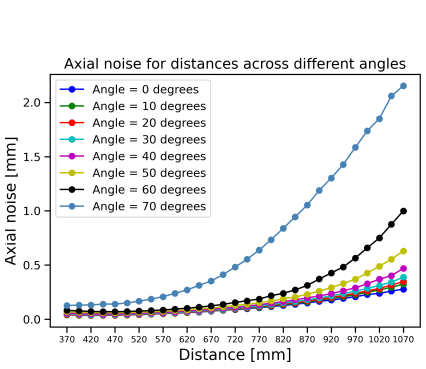
3.3. Axial Noise

Axial noise refers to depth measurement errors that vary along the sensor's viewing axis, characterized by deviation of points from a plane fitted to the scan of the planar target.

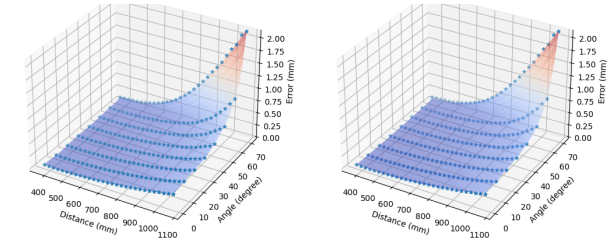
Plane fitting: we initially extract a clean surface of the calibration board from the scan, excluding boundaries to eliminate lateral noise effects and other unwanted effects such as specular reflections. Subsequently, a plane is fitted to the clean extracted surface and the error is estimated as a deviation between the plane fitted and the surface points. This process is repeated for all distances covered by the captured point clouds, across angles ranging from 0° to 70° .

Given a set of N points $\mathcal{P} = \{\mathbf{p_i}\}_{i=1}^{N}$ from the extracted clean point cloud, where $\mathbf{p_i} = (x_i, y_i, z_i) \in \mathbb{R}^3$, at a distance z and angle θ with respect to the Zivid camera, the plane fitting problem aims to find the coefficients $X = \{a, b, c, d\}$, such that the plane equation $ax + by + cz + d = 0$ best fits the given points in the least square sense.

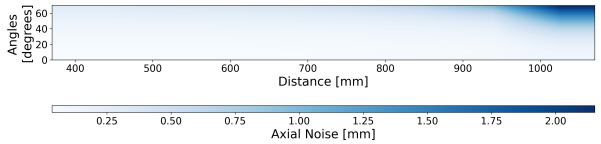
Fitted axial noise model: the axial noise for a single scan can then be calculated as the root mean square distance between the fitted plane coefficients and the points. Figure 4a shows axial noise values as a function of depth and angles. The axial noise is more pronounced at larger distances for an angle of 50° or greater. The noise exhibits a non-linear trend, increasing as both the surface distance and surface angle increase.



(**a**) Axial noise values as a function of distance and angle between the camera and the planar target



(**b**) polynomial $f_1(z,\theta)$ (**c**) exp. + polynomial $f_2(z,\theta)$



(**d**) Interpolated image encoding for axial noise values

Figure 4. Axial noise modeling. (**b**,**c**) Fitted surface models for axial noise with a (**b**) 7th order bivariate polynomial, (**c**) bivariate exponential function plus a 2nd order bivariate polynomial, and (**d**) bilinearly interpolated image encoding for axial noise values corresponding to (**a**). The axis labels and titles are provided for illustrative purposes.

To capture this non-linear trend, we apply three distinct methods to fit a function to the noise curves in Figure 4a, and we select the one most convenient for simplified integration into 3D reconstruction algorithms.

Polynomial fitting: we fit surface models for the axial noise data displayed in Figure 4a. The model used in Figure 4b is represented as $f_1(z,\theta) = \sum_{i=0}^{7}\sum_{j=0}^{7-i} a_{ij} z^i \theta^j$ for an order 7 bivariate polynomial, with 36 coefficients on a_{ij}.

Polynomial+exponential fitting: to reduce the number of parameters fitted, we choose to mix both polynomials and exponential terms in the fitting to handle distortions and distance variations. The model used in Figure 4c is $f_2(z,\theta) = Se^{Az^2+Bz\theta+C\theta^2+Dz+E\theta} + az^2 + bz\theta + c\theta^2 + dz + e\theta + f$, with $\{S, A, B, C, D, E, a, b, c, d, e, f\}$ being the coefficients. There are still 12 coefficients for $f_2(z,\theta)$. The values of all coefficients for both models are provided in the supplementary material in Tables S1 and S2.

Look-up image: due to the complexity of higher-order coefficients in both fitted surfaces, we generate an alternate look-up image, as shown in Figure 4d, whose pixels store values of the axial noise obtained in Figure 4a. Using bilinear interpolation, noise values are interpolated for any combination of distance and surface angle. This resulting axial noise image can easily be queried in downstream applications, such as KinectFusion [10], to obtain the noise value corresponding to a given distance and angle to achieve better 3D reconstruction. Table S3 in the Supplementary Material lists the axial noise values corresponding to distance and angle as shown in Figure 4.

3.4. Lateral Noise

Lateral noise refers to depth measurement errors that occur along the directions perpendicular to the camera axis that mainly looks at edge pixels of the planar target scan.

Line fitting: Figure 5 explains the steps involved in lateral noise modeling from depth maps. Firstly, we remove the background pixels of the planar target. Then, we find the position of the top edge pixels of the planar target in the depth map. We limit the number of pixels to the middle 200 points of the top edge to avoid the effects of optical distortion.

Subsequently, we perform a line fitting on the x and y coordinates of the extracted edge pixel points. We then calculate σ_L as the root mean square distances between the pixel coordinates of the edge and the fitted line.

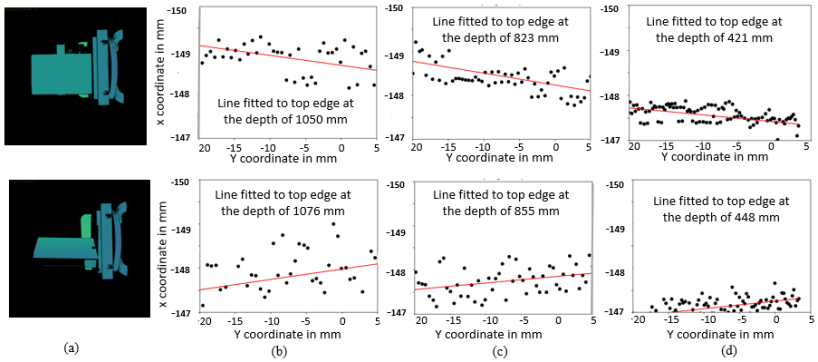


Figure 5. Pre-processing steps for lateral noise modeling. (**a**) Depth map with segmented edge of the planar target (marked in red color); Line fitted to pixels corresponding to the edge: at *far* distance (**b**), at *medium* distance (**c**), and at *near* distance (**d**). Rows indicate the angle of rotation of target, varying from 0° (**top**) to 60° (**bottom**).

Fitted lateral noise model: this procedure is repeated for all scans obtained across distances and angles ranging from 0° to 80°. Ideally, a single line should adequately represent all edge pixel coordinates, but due to lateral noise effects this is not always the case. Figure 5 shows that the spread of the points around the fitted line decreases with decreasing distances. We observe that the lateral noise is independent of the surface angles from 0° to 60°, but increases significantly for 70° to 80°, indicating the reduced reliability of the measurements at extreme surface angles. Figure 6 shows the fit of the linear model to the lateral noise, σ_L, as a function of surface distance.

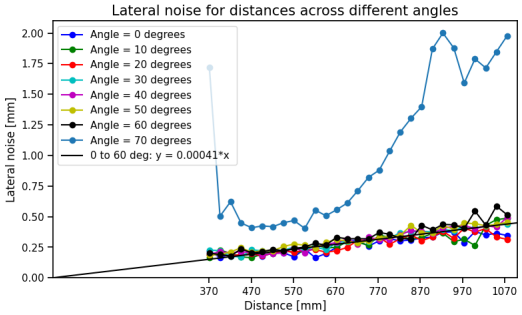


Figure 6. Lateral noise against different distances and angles.

3.5. Other Effects to Sensor Noise

In this section, we explore the impact of varying lighting conditions, exposure times, and capture settings on axial and lateral noise as functions of distance. Figure 7 (left) shows that both axial and lateral noise are independent of the background lighting, since the light emitted by the structured light sensor is much stronger compared with the ambient background lighting. Figure 7 (middle) shows that the quality of the scans can be improved by increasing the number of captures, each with different aperture settings, to capture with a high dynamic range. Figure 7 (right) shows that increased exposure time leads to a decrease in lateral noise due to the smoothing effect at the depth edges. However, for axial noise, the quality of the depth images becomes degraded for extended exposure time due to sensor saturation.

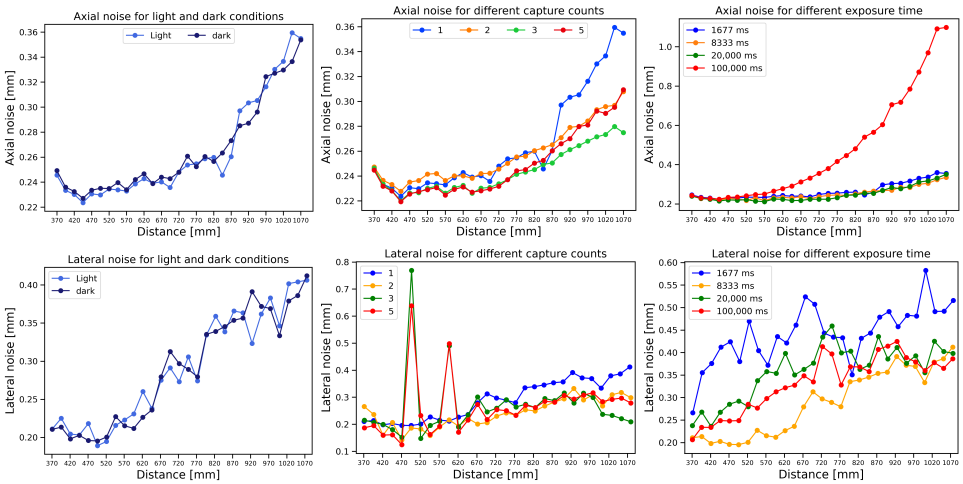


Figure 7. Variations in axial (**top**) and lateral noise (**bottom**) due to lighting conditions (**left**), number of captures (**middle**), and exposure time (**right**) as a function of distances at fixed 0°.

4. Applying Noise Model to Improve 3D Reconstruction

To validate the effectiveness of our noise model, we capture a dataset using a camera mounted on a robot arm, with objects placed on a turntable. We then integrate our noise model into both traditional and implicit neural-based 3D reconstruction methods to improve the quality of both 3D reconstruction and pose estimation. We primarily seek to demonstrate how our modeled noise improves the quality of reconstruction using traditional KinectFusion, which is known for its reliability, compared with modern neural-based reconstruction methods [9].

4.1. Datasets

In our experiment, we scan a variety of test objects mounted on a motorized turntable, using a Zivid camera integrated with a UR5e robotic arm. The test objects are cultural artifacts (Shiva and Ganesh), toy models (Dino and Dragon), and precision-engineered components (Gripper and Controller), as depicted in Figure 8c. The scanning process involves rotating each object on the turntable with an incremental step of 1°. At each step, the Zivid camera captures an RGB-D image, resulting in a total of 360 images per object along a circular path. This precise rotational increment facilitates the effectiveness of camera tracking algorithms, which often operate under the assumption of small angular changes. Consequently, the generated dataset provides comprehensive ground-truth information, enabling quantitative comparisons for each scanned object.

In order to obtain a baseline for 3D reconstruction for the dataset, we use traditional ICP registration for pose estimation and fusing multiple point clouds, followed by Poisson reconstruction for generating a mesh. Point-cloud data are filtered using a Euclidean cluster extraction algorithm, and then pair-wise data registration is performed using ICP to obtain the transformation parameters between two consecutive camera poses, followed by pose graph optimization. A subset of the whole point cloud data is chosen for merging to reduce the data size. Finally, Poisson 3D surface reconstructions are performed on the data obtained. Figure 9 shows the 3D surface reconstructions obtained for the Shiva and Gripper datasets.

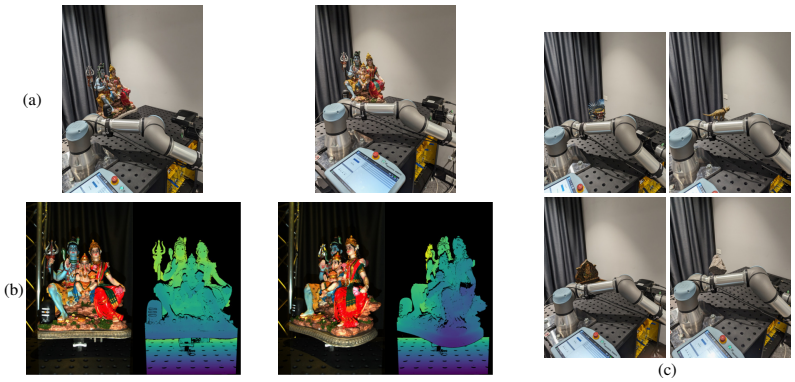


Figure 8. The experimental setup for dataset collection. From top to bottom: (**a**) complete setup with a robot arm, camera, and object placed in a rotating table for 0° and 45° rotation of the object, (**b**) corresponding RGB and scaled depth images captured, and (**c**) other object sequences captured—Dragon, Ganesh, Rock, and Dino.



(**a**) Shiva (**b**) Gripper

Figure 9. Poisson 3D surface reconstructions with point cloud data merged based on pair-wise ICP registration followed by pose graph optimization for (**a**) Shiva and (**b**) Gripper.

4.2. Improving KinectFusion Using Our Noise Model

Following a similar approach as [5], we integrate our noise model into the KinectFusion algorithm [10] to improve the approximation of the truncated sign distance function (TSDF) for each voxel on the voxel grid. KinectFusion iteratively fuses depth data from an RGB-D camera to a voxel grid structure stored on the GPU. Within this grid, surface details are implicitly encoded as signed distances, truncated to a predefined layer containing the expected surface of an object. The integration of new depth values into the voxel grid employs a weighted average scheme. The global pose of the camera is tracked using a linearized ICP algorithm (using small angle assumption) [42] by aligning the current frame with the accumulated model. The implicit surface can be extracted by either ray-casting the voxel grid or triangulating a mesh using marching cubes. A 2D pixel coordinate on the depth map is denoted as $u = (x, y)$. $D_i(u)$ is the depth value at pixel u retrieved at the ith frame. With an intrinsic calibration matrix, K, which is typically a 3×3 matrix containing the camera's focal lengths and optical center, a 3D vertex corresponding to the pixel u is $v_i(u) = D_i(u)K^{-1}[u, 1]$. Here, K^{-1} is the inverse of the intrinsic matrix K, and $[u, 1]$ is the homogeneous coordinate representation of the pixel u. D_i therefore produces a single vertex map, V_i, which represents the 3D points for all pixels at the ith frame. The camera pose in the ith frame is represented as $T_i = [R_i, c_i]$, where R_i is the rotation matrix and c_i is the translation vector, which describes the position and orientation of the camera relative to the world coordinate system. The position of the vertex in the global coordinates is then expressed as $v_{g_i} = T_i \cdot v_i$, where v_{g_i} is the transformed vertex in the global coordinates, and $T_i \cdot v_i$ applies the rotation and translation to the 3D point v_i. The normal vector $n_i(p_k)$ at a point p_k in 3D space represents the direction perpendicular to the surface at that point. It is

used to calculate the angle, θ, between the z-axis and the normal, allowing surface data to be weighted based on the orientation of the depth measurement, with smaller angles indicating more reliable contributions to the updated TSDF.

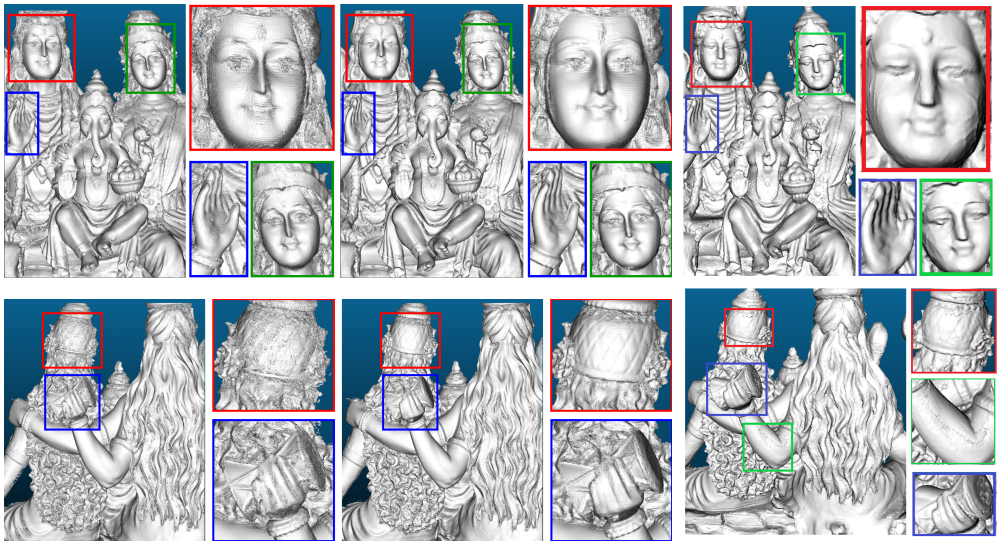
Algorithm 1 states the KinectFusion algorithm in which our noise model has been incorporated for enhancing the TSDF values leading to improved depth fusion. The TSDF calculation at line 26 of the algorithm is designed to generate values within the range of $[-1, 1]$. In contrast to [10], where a uniform weight, w, is assigned to all depth measurements regardless of their noise levels for reconstruction updates, we address this issue by assigning weights to depth measurements based on their noise quality and distance from the camera, as shown in Line 27 of Algorithm 1. The exponential term represents the weight of the noise distribution. $\sigma_z(z_{\min})$ represents the axial noise value at the minimum depth (z_{min}) used in the noise modeling process. For a detailed explanation of the steps of the KinectFusion [10] algorithm, refer to [5] and the references therein.

Algorithm 1 TSDF integration considering both axial and lateral noise using bilinear interpolation

1: **Input:**
2: Depth map D_i for frame i
3: Camera pose T_i consisting of rotation matrix R_i and translation vector c_i
4: Noise models: σ_Z, σ_L
5: Angle threshold θ_{max}
6: Size, location and resolution of 3D volume V containing the object to be reconstructed
7: **Output:**
8: Updated TSDF $tsdf_i$ for the current frame
9: **for** each voxel g on x-y slice of volume V in parallel **do**
10: **while** sweep along z-axis of volume **do**
11: $v_g \leftarrow$ convert g from grid to global 3D position
12: $v \leftarrow T_i^{-1} v_g$
13: $p \leftarrow$ perspective project vertex v
14: **if** $p \in$ depth map D_i and $D_i(p) > 0$ **then**
15: $tsdf_i \leftarrow tsdf_{i-1}$
16: $w_i \leftarrow w_{i-1}$
17: **for** each p_k in 3×3 area around p **do**
18: $\Delta u \leftarrow ||p - p_k||$ with sub-pixel accuracy
19: $\theta \leftarrow$ angle(z-axis, $n_i(p_k)$)
20: $\Delta z \leftarrow ||D_i(p) - D_i(p_k)||$
21: **if** $D_i(p_k) > 0$ and $\theta < \theta_{max}$ **then**
22: $\sigma_Z \leftarrow$ Look up from noise image (Figure 4) against $D_i(p_k)$ and θ
23: $\sigma_L \leftarrow \begin{cases} f_x \times 0.00041 & \text{for } \theta \leq 60 \\ f_x \times 0.0043 & \text{for } \theta > 60 \end{cases}$ (Figure 6)
24: $sdf_k \leftarrow ||c_i - v|| - D_i(p_k)$
25: **if** $sdf_k > -6\sigma_z$ and $\Delta z < 3\sigma_z$ **then**
26: $tsdf_k \leftarrow \text{sgn}(sdf_k) \sqrt{1 - e^{-\frac{2\text{sdf}_k^2}{\pi\sigma_Z^2}}}$
27: $w_k \leftarrow \frac{\sigma_z(z_{\min})}{\sigma_Z} \frac{z_{\min}^2}{D_i^2(p_k)} e^{-\frac{\Delta u^2}{2\sigma_L^2} - \frac{\Delta z^2}{2\sigma_z^2}}$
28: $tsdf_i \leftarrow \frac{tsdf_i w_i + tsdf_k w_k}{w_i + w_k}$
29: $w_i \leftarrow \min(\text{max weight}, w_i + w_k)$
30: **end if**
31: **end if**
32: **end for**
33: **end if**
34: **end while**
35: **end for**

Algorithm S2, attached in the supplementary material, provides an efficient implementation of Algorithm 1. By skipping lateral noise, which requires iteration over the 3×3 neighborhood, this version significantly speeds up runtime ($\approx$9 times) with a small performance drop (results in the Supplementary Materials).

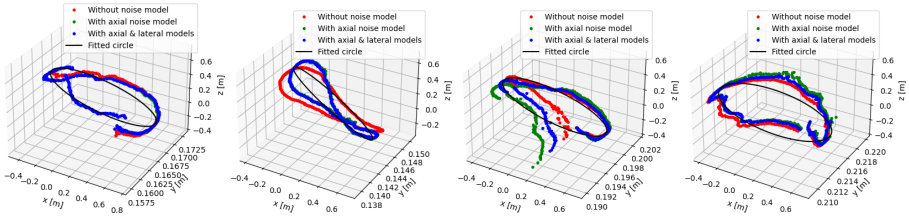
Figure 10b illustrates the substantial improvement in reconstruction quality with the KinectFusion algorithm using our noise model compared with the results obtained without noise filtering (Figure 10a) and with a more traditional 3D reconstruction pipeline (Figure 10c), as explained in Section 4.1.



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

Figure 10. The qualitative results from the Shiva dataset, both top-front and bottom-back views, highlighting improved reconstruction quality (**b**), in the colored boxes, when noisy depth measurements are filtered using our noise model in the KinectFusion algorithm. This improvement is particularly noticeable compared with the reconstruction without depth filtering (**a**) and against the traditional reconstruction pipeline (**c**) discussed in Section 4.1.

This depth filtering using our noise also improves tracking, as shown in Figure 11, where we observe a reduced drift in the trajectory when the depth is filtered. The decreased drift for the noisy depth filtering case is evident from the minimal difference between the start and end poses of the camera for 360° rotations, which is significantly large when noisy depth measurements are not filtered. The corresponding RMSE between the ground-truth poses and the trajectories obtained with and without noise filtering is presented in Table 1. The ground-truth poses were obtained from the calibrated rotating table and the robot arm setup, which were further refined by fitting a circular plane on the rotating axis.



(**a**) Controller (**b**) Rock (**c**) Dragon (**d**) Dino

Figure 11. Comparison of trajectories for different objects with and without noise filtering. Results demonstrate superior trajectories with noise filtering (green and blue) as compared with those without noise (red). Fitted circle to the trajectory with axial and lateral noise models (blue). The axes are not of the same scale.

Table 1 provides drift measurements (both distance and angle) as a quantitative comparison between different reconstruction conditions (without noise, with axial noise, and with axial and later noise models) for five datasets. The reconstruction with noise model gains significant improvements for Shiva, Controller and Rock, although it slightly underperforms reconstruction without noise models for Dragon and Dino. Most of the time, reconstruction with just axial noise performs close to that with both axial and lateral noise models, except for very smooth objects, and at a significant speed-up of approximately 9-fold.

Table 1. Drift values for translation and rotation (m, degrees) of first and last estimated camera poses of different objects.

Objects	Without Noise	Axial Noise	Axial-Lateral Noise
Shiva	(0.00613, 0.42107)	(0.00184, 0.12915)	(**0.00178, 0.12375**)
Controller	(1.13467, 33.417)	(1.08114, 48.397)	(**0.00579, 0.58854**)
Rock	(0.00240, 0.31135)	(0.00086, 0.13408)	(**0.00080, 0.10139**)
Dragon	(**0.00250, 0.41372**)	(0.01116, 1.7199)	(0.00820, 1.2446)
Dino	(**0.01085, 1.159**)	(0.01121, 1.1883)	(0.01131, 1.2041)

Table 2 provides a quantitative evaluation of the Gripper mesh across four cases, reconstructed using ground-truth poses, compared with the ground-truth mesh. It emphasizes the critical role of depth map filtering using our empirically derived noise models in applications demanding submillimeter-level accuracy. The corresponding precision and recall plots across various distance thresholds are detailed in Figure S7 of the Supplementary Materials.

Table 2. Comparison of Gripper mesh with a ground-truth mesh in terms of precision (Prec.), recall (Rec.), and F-score (F) at a distance threshold of 0.1 mm.

Baseline			**Without Noise**			**Axial Noise**			**Axial-Lateral Noise**		
Prec.↑	Rec.↑	F↑	Prec.↑	Rec.↑	F↑	Prec.↑	Rec.↑	F↑	Prec.↑	Rec.↑	F↑
0.154	0.435	0.228	0.361	0.479	0.412	0.382	**0.485**	0.427	**0.385**	0.482	**0.428**

4.3. Improving Neural Implicit SLAM Using Our Noise Model

We are able to integrate our noise model in NeRF-based SLAM systems such as Point-SLAM [9], as shown in Figure 12, by incorporating uncertainty weighting and ground-truth depth refinement based on our noise model. The ground-truth depth refinement is performed by calculating the expected depth value from a neighborhood of 3×3 pixels using a Gaussian distribution that incorporates both axial and lateral noise. We use $\sigma_z(z_{\min})/\sigma_z(z)$ as the uncertainty weight, where $\sigma_z(z_{\min})$ represents the axial noise at the minimum depth value. Moreover, we incorporate this axial noise as a dynamic lower bound in the point search radius of Point-SLAM. This dynamic lower bound is critical for ensuring that the search for neighboring points spans a sufficient spatial range, especially for distant points.

Our noise modeling approach focuses primarily on high-resolution Zivid scans. Integrating our noise model significantly improves results when the reconstruction is performed on a very fine scale (0.2 mm or finer). This aligns well with traditional multi-view stereo and classical point-based reconstruction methods. Difficulties are encountered when applying our noise model to neural implicit approaches. When using Gaussian Splatting-based SLAM (such as Splat-TAM [43]), memory limitations restrict the resolution at which our noise model can be effectively utilized. When using NeRF-based SLAM (such as Point-SLAM [9]), we encounter computational performance bottlenecks. Incorporating high-resolution noise corrections efficiently into these methods will require further research.



(a) (b) (c)

Figure 12. Qualitative results of integrating our noise model in Point-SLAM with the Shiva dataset. Highlighted boxes show improved reconstruction—Point-SLAM baseline (**a**), noisy depth filtering using our noise model (**b**), and depth loss term using our noise model (**c**). Please zoom in the red box for better visualization.

5. Conclusions

In this paper, we empirically derive a noise model for the RGB-D Zivid sensor that quantifies the axial and lateral noise components as a function of distance and surface angle. Experimental results demonstrate quantitatively and qualitatively that our derived noise model leads to more accurate 3D reconstructions. Integrating this model into 3D reconstruction pipelines, through bilinear interpolation or fitted functions, yields significant improvements in both pose estimation and reconstruction accuracy. Therefore, the use of our method can be recommended in advanced manufacturing and other applications that require both high resolution and reduced measurement uncertainty. Additionally, we provide a publicly available dataset of high-resolution RGB-D scans to facilitate future applied research.

A limitation of our current approach is that it is based on scanning calibration targets to accurately estimate sensor noise characteristics. A promising direction for future research is the development of self-supervised or data-driven methods that can infer noise characteristics directly from captured scenes without requiring explicit calibration setups. Another limitation is that, while we have empirically modeled the impact of environmental factors such as temperature and background lighting on sensor noise, we have not incorporated these effects into our reconstruction pipeline. Future work can focus on dynamically integrating these models to improve the robustness of reconstruction under varying environmental conditions. Additionally, although our noise model has demonstrated an improvement in 3D reconstruction, we have not systematically studied how different reconstruction paradigms, such as traditional multi-view stereo versus implicit neural representations, are affected by noise. A potential research direction is to analyze how our noise model can be uniformly applied across these diverse approaches to ensure both consistency and robustness in reconstruction. Finally, our current noise model focuses primarily on estimating the standard deviation of noise. However, sensor noise can also introduce systematic biases that affect the accuracy of the reconstruction. Future research should explore methods to estimate and compensate for these bias components in RGB-D sensor data, enabling more precise 3D reconstructions.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/s25030950/s1, Figure S1: Comparison between 3D reconstruction obtained using a full noise model incorporating both axial and lateral noise components (left), axial noise using Algorithm 1 of the main text (middle), and Algorithm S1 (right). We can see slight compromise in the quality with the later (right) but at a roughly nine fold improved speed; Figure S2: Comparison of reconstruction quality using ground truth camera poses, without noise (left), with axial noise only (middle), and with both axial and lateral noise (right). Our noise model integrated

with TSDF demonstrates superior reconstruction quality as shown in zoomed in images at the bottom of each figure. When comparing the effectiveness of noise filtering using just axial noise versus utilizing both axial and lateral noise, we observe that incorporating both components yields higher resolution for fine details in the reconstructed depth data; Figure S3: Comparison of the quality of reconstructions (top) and trajectories (bottom) with unfiltered depth maps on the left and filtered results with axial noise in the middle and both axial and lateral noise on the right. Depth map filtering using our noise model improves both the reconstruction quality and camera pose tracking. Note that the axes of trajectories are not of the same scale; Figure S4: Reconstruction results (top to bottom) for Controller, Rock, Dragon, and Dino. From left to right we have without noise filtering, with axial noise filtering, and with both axial and lateral noise filtering. On these relatively smooth objects, the difference in reconstruction quality between filtered and unfiltered versions is minimal. This is attributed to the lack of high resolution geometric features in these objects which our noise model significantly enhances; Figure S5: Poor quality of the gripper's mesh using the Kinect noise model compared to the Zivid noise model in Figure S10; Figure S6: Poor quality of the gripper's mesh using the Gaussian filtering compared to the Zivid noise model in Figure S10; Figure S7: Precision and recall metrics across different distance thresholds for four meshes: (a) baseline method, (b) without noise filtering, (c) with axial noise filtering, and (d) with both axial and lateral noise filtering; Figure S8: Precision and Recall meshes along with their histograms for baseline method; Figure S9: Precision and Recall meshes along with their histograms without depth filtering; Figure S10:Precision and Recall meshes along with their histograms for axial noise filtering case; Figure S11: Precision and Recall meshes along with their histograms for axial and lateral noise filtering case; Table S1: Coefficients for $f_1(z, \theta)$; Table S2: Coefficients for $f_2(z, \theta)$; Table S3: Axial noise values corresponding to Figure 4a of the main text; Algorithm S1: TSDF integration with axial noise only.

Author Contributions: Conceptualization, F.A.M., S.M., C.N. and C.S.; Data curation, F.A.M. and S.M.; Formal analysis, F.A.M., S.M., C.N., C.S. and J.T.; Funding acquisition, S.D., C.N. and L.P.; Investigation, F.A.M., S.M., C.N. and C.S.; Methodology, F.A.M., S.M., C.N. and C.S.; Project administration, C.N., S.D. and L.P.; Resources, C.N.; Software, F.A.M., S.M., C.N., C.S., J.T. and S.W.; Supervision, C.N., D.H., S.D. and L.P.; Validation, F.A.M., S.M., C.N., C.S., J.T. and S.W.; Visualization, F.A.M., S.M., C.N., C.S. and J.T.; Writing—original draft, F.A.M., S.M. and J.T.; Writing—review and editing, F.A.M., S.M., C.N., C.S., R.T., S.D. and L.P. All authors have read and agreed to the published version of the manuscript.

Funding: Future Digital Manufacturing Fund, Julius Career Award, CSIRO CERC Fellowship.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article and Supplementary Materials.

Acknowledgments: We would like to express our sincere gratitude to Lars Andersson (CSIRO Data61) for his invaluable assistance in resolving hardware setup issues and his significant contributions to various aspects of the project.

Conflicts of Interest: The authors declare that there are no conflicts of interest regarding the publication of this paper. However, there are specific conflicts related to affiliations with the Australian National University, and CSIRO Australia.

References

1. Han, J.; Shao, L.; Xu, D.; Shotton, J. Enhanced Computer Vision with Microsoft Kinect Sensor: A Review. *IEEE Trans. Cybern.* **2013**, *43*, 1318–1334. [PubMed]
2. Haider, A.; Hel-Or, H. What Can We Learn from Depth Camera Sensor Noise? *Sensors* **2022**, *22*, 5448. [CrossRef] [PubMed]
3. Jing, C.; Potgieter, J.; Noble, F.; Wang, R. A comparison and analysis of RGB-D cameras' depth performance for robotics application. In Proceedings of the 2017 24th International Conference on Mechatronics and Machine Vision in Practice (M2VIP), Auckland, New Zealand, 21–23 November 2017; pp. 1–6.

4. Tölgyessy, M.; Dekan, M.; Chovanec, L.; Hubinskỳ, P. Evaluation of the Azure Kinect and its comparison to Kinect v1 and Kinect v2. *Sensors* **2021**, *21*, 413. [CrossRef]
5. Nguyen, C.V.; Izadi, S.; Lovell, D. Modeling Kinect sensor noise for improved 3D reconstruction and tracking. In Proceedings of the 2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization & Transmission, Zurich, Switzerland, 13–15 October 2012; pp. 524–530.
6. Fankhauser, P.; Bloesch, M.; Rodriguez, D.; Kaestner, R.; Hutter, M.; Siegwart, R. Kinect v2 for mobile robot navigation: Evaluation and modeling. In Proceedings of the 2015 International Conference on Advanced Robotics (ICAR), Istanbul, Turkey, 27–31 July 2015; pp. 388–394.
7. Mallick, T.; Das, P.P.; Majumdar, A.K. Characterizations of noise in Kinect depth images: A review. *IEEE Sensors J.* **2014**, *14*, 1731–1740. [CrossRef]
8. Khoshelham, K.; Elberink, S.O. Accuracy and Resolution of Kinect Depth Data for Indoor Mapping Applications. *Sensors* **2012**, *12*, 1437–1454. [CrossRef]
9. Sandström, E.; Li, Y.; Van Gool, L.; Oswald, M.R. Point-SLAM: Dense neural point cloud-based SLAM. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 2–6 October 2023; pp. 18433–18444.
10. Izadi, S.; Kim, D.; Hilliges, O.; Molyneaux, D.; Newcombe, R.; Kohli, P.; Shotton, J.; Hodges, S.; Freeman, D.; Davison, A.; et al. KinectFusion: Real-time 3D reconstruction and interaction using a moving depth camera. In Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology, Santa Barbara, CA, USA, 16–19 October 2011; pp. 559–568.
11. Choo, B.; DeVore, M.D.; Beling, P.A. Statistical models of horizontal and vertical stochastic noise for the Microsoft Kinect™ sensor. In Proceedings of the IECON 2014—40th Annual Conference of the IEEE Industrial Electronics Society, Dallas, TX, USA, 29 October–1 November 2014; pp. 2624–2630.
12. Halmetschlager-Funek, G.; Suchi, M.; Kampel, M.; Vincze, M. An empirical evaluation of ten depth cameras: Bias, precision, lateral noise, different lighting conditions and materials, and multiple sensor setups in indoor environments. *IEEE Robot. Autom. Mag.* **2018**, *26*, 67–77. [CrossRef]
13. Dong, W.; Wang, Q.; Wang, X.; Zha, H. PSDF fusion: Probabilistic signed distance function for on-the-fly 3D data fusion and scene reconstruction. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 701–717.
14. Proenca, P.F.; Gao, Y. Probabilistic RGB-D odometry based on points, lines and planes under depth uncertainty. *Robot. Auton. Syst.* **2018**, *104*, 25–39. [CrossRef]
15. Dryanovski, I.; Valenti, R.G.; Xiao, J. Fast visual odometry and mapping from RGB-D data. In Proceedings of the International Conference on Robotics and Automation (ICRA), Karlsruhe, Germany, 6–10 May 2013; pp. 2305–2310.
16. Gutierrez-Gomez, D.; Mayol-Cuevas, W.; Guerrero, J.J. Dense RGB-D visual odometry using inverse depth. *Robot. Auton. Syst.* **2016**, *75*, 571–583. [CrossRef]
17. Wasenmüller, O.; Ansari, M.D.; Stricker, D. Dense noise aware SLAM for tof RGB-D cameras. In Proceedings of the Asian Conference on Computer Vision Workshop (ACCV Workshop), Taipei, Taiwan, 20–24 November 2016.
18. Yamaguchi, T.; Emaru, T.; Kobayashi, Y.; Ravankar, A.A. 3D map-building from RGB-D data considering noise characteristics of Kinect. In Proceedings of the 2016 IEEE/SICE International Symposium on System Integration (SII), Sapporo, Japan, 13–15 December 2016; pp. 379–384.
19. Rosinol, A.; Leonard, J.J.; Carlone, L. Probabilistic Volumetric Fusion for Dense Monocular SLAM. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 2–7 January 2023; pp. 3097–3105.
20. Brandao, M.; Figueiredo, R.; Takagi, K.; Bernardino, A.; Hashimoto, K.; Takanishi, A. Placing and scheduling many depth sensors for wide coverage and efficient mapping in versatile legged robots. *Int. J. Robot. Res.* **2020**, *39*, 431–460. [CrossRef]
21. Lu, Y.; Song, D. Robust RGB-D odometry using point and line features. In Proceedings of the International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015.
22. Smisek, J.; Jancosek, M.; Pajdla, T. 3D with Kinect. In *Consumer Depth Cameras for Computer Vision*; Springer: Berlin/Heidelberg, Germany; pp. 3–25.
23. Iversen, T.M.; Kraft, D. Generation of synthetic Kinect depth images based on empirical noise model. *Electron. Lett.* **2017**, *53*, 856–858. [CrossRef]
24. Glover, J.; Popovic, S. Bingham procrustean alignment for object detection in clutter. In Proceedings of the 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems, Tokyo, Japan, 3–7 November 2013; pp. 2158–2165.
25. Saulnier, K.; Atanasov, N.; Pappas, G.J.; Kumar, V. Information theoretic active exploration in signed distance fields. In Proceedings of the 2020 IEEE International Conference on Robotics and Automation (ICRA), Paris, France, 31 May–31 August 2020; pp. 4080–4085.
26. Kerl, C.; Sturm, J.; Cremers, D. Robust odometry estimation for RGB-D cameras. In Proceedings of the International Conference on Robotics and Automation (ICRA), Karlsruhe, Germany, 6–10 May 2013; pp. 3748–3754.

27. Kerl, C.; Sturm, J.; Cremers, D. Dense visual SLAM for RGB-D cameras. In Proceedings of the International Conference on Intelligent Robots and Systems (IROS), Tokyo, Japan, 3–7 November 2013; pp. 2100–2106.
28. Dhawale, A.; Michael, N. Efficient Parametric Multi-Fidelity Surface Mapping. In Proceedings of the Robotics: Science and Systems, Virtually, Los Angeles, CA, USA, 12–16 July 2020.
29. Oleynikova, H.; Taylor, Z.; Fehr, M.; Siegwart, R.; Nieto, J. Voxblox: Incremental 3D Euclidean Signed Distance Fields for On-Board MAV Planning. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Vancouver, BC, Canada, 24–28 September 2017; pp. 1366–1373.
30. Ran, Y.; Zeng, J.; He, S.; Li, L.; Chen, Y.; Lee, G.; Chen, J.; Ye, Q. NeurAR: Neural Uncertainty for Autonomous 3D Reconstruction With Implicit Neural Representations. *IEEE Robot. Autom. Lett.* **2023**, *8*, 1125–1132. [CrossRef]
31. Osvaldov'a, K.; Gajdovsech, L.; Kocur, V.; Madaras, M. Enhancement of 3D Camera Synthetic Training Data with Noise Models. *arXiv* **2024**, arXiv:2402.16514.
32. Cai, Y.; Plozza, D.; Marty, S.; Joseph, P.; Magno, M. Noise Analysis and Modeling of the PMD Flexx2 Depth Camera for Robotic Applications. In Proceedings of the 2024 IEEE International Conference on Omni-layer Intelligent Systems (COINS), London, UK, 29–31 July 2024; pp. 1–6.
33. Rustler, L.; Volprecht, V.; Hoffmann, M. Empirical Comparison of Four Stereoscopic Depth Sensing Cameras for Robotics Applications. arXiv **2025**, arXiv:2501.07421. [CrossRef]
34. Burger, L.; Sharan, L.; Karl, R.; Wang, C.; Karck, M.; De Simone, R.; Wolf, I.; Romano, G.; Engel-hardt, S. Comparative evaluation of three commercially available markerless depth sensors for close-range use in surgical simulation. *Int. J. Comput. Assist. Radiol. Surg.* **2023**, *18*, 1109–1118. [CrossRef] [PubMed]
35. Ramli, I.S.; OK Rahmat, R.W.; Ng, S.B. Enhancement of Depth Value Approximation for 3D Image-Based Modelling using Noise Filtering and Inverse Perspective Mapping Techniques for Complex Object. *J. Comput. Res. Innov.* **2023**, *8*, 246–264.
36. Nießner, M.; Zollhöfer, M.; Izadi, S.; Stamminger, M. Real-time 3D reconstruction at scale using voxel hashing. *ACM Trans. Graph. (Tog)* **2013**, *32*, 169. [CrossRef]
37. Newcombe, R.A.; Fox, D.; Seitz, S.M. DynamicFusion: Reconstruction and tracking of non-rigid scenes in real-time. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 343–352.
38. Runz, M.; Buffier, M.; Agapito, L. MaskFusion: Real-time recognition, tracking and reconstruction of multiple moving objects. In Proceedings of the 2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Munich, Germany, 16–20 October 2018; pp. 10–20.
39. Dai, A.; Nießner, M.; Zollhöfer, M.; Izadi, S.; Theobalt, C. BundleFusion: Real-time globally consistent 3D reconstruction using on-the-fly surface reintegration. *ACM Trans. Graph. (ToG)* **2017**, *36*, 1. [CrossRef]
40. Zhang, Y.; Tosi, F.; Mattoccia, S.; Poggi, M. Go-SLAM: Global optimization for consistent 3D instant reconstruction. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 1–6 October 2023; pp. 3727–3737.
41. Xu, Q.; Xu, Z.; Philip, J.; Bi, S.; Shu, Z.; Sunkavalli, K.; Neumann, U. Point-NeRF: Point-based 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. 5438–5448.
42. Besl, P.J.; McKay, N.D. A method for registration of 3-D shapes. *IEEE Trans. Pattern Anal. Mach. Intell.* **1992**, *14*, 239–256. [CrossRef]
43. Keetha, N.; Karhade, J.; Jatavallabhula, K.M.; Yang, G.; Scherer, S.; Ramanan, D.; Luiten, J. SplaTAM: Splat, Track and Map 3D Gaussians for Dense RGB-D SLAM. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 11–15 June 2024; pp. 21357–21366.

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.