

Article

# Maximum Sum of Evidence—An Evidence-Based Solution to Object Pose Estimation in Point Cloud Data

Tyson Phillips , Tim D'Adamo  and Peter McAree 

School of Mechanical and Mining Engineering, The University of Queensland, Brisbane, QLD 4072, Australia; t.dadamo@uq.edu.au (T.D.); p.mcaree@uq.edu.au (P.M.)

\* Correspondence: t.phillips1@uq.edu.au; Tel.: +61-733-653-587

**Abstract:** The capability to estimate the pose of known geometry from point cloud data is a frequently arising requirement in robotics and automation applications. This problem is directly addressed by Iterative Closest Point (ICP), however, this method has several limitations and lacks robustness. This paper makes the case for an alternative method that seeks to find the most likely solution based on available evidence. Specifically, an evidence-based metric is described that seeks to find the pose of the object that would maximise the conditional likelihood of reproducing the observed range measurements. A seedless search heuristic is also provided to find the most likely pose estimate in light of these measurements. The method is demonstrated to provide for pose estimation (2D and 3D shape poses as well as joint-space searches), object identification/classification, and platform localisation. Furthermore, the method is shown to be robust in cluttered or non-segmented point cloud data as well as being robust to measurement uncertainty and extrinsic sensor calibration.

**Keywords:** pose estimation; object classification; localisation; perception; LiDAR



**Citation:** Phillips, T.; D'Adamo, T.; McAree, R. Maximum Sum of Evidence—An Evidence-Based Solution to Object Pose Estimation in Point Cloud Data. *Sensors* **2021**, *21*, 6473. <https://doi.org/10.3390/s21196473>

Academic Editor: Antonio M. Pascoal

Received: 24 August 2021

Accepted: 27 September 2021

Published: 28 September 2021

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



**Copyright:** © 2021 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

The capability for robotic agents to establish beliefs about their environment is fundamental. These beliefs often relate to objects whose geometry is known but their presence and location needs to be established. Three problems typically arise. Given point cloud data: (i) what is the most likely pose of an object of known geometry believed to be present in the data? (ii) which object, from among a set whose geometries are known, is most likely to be present in the data? (iii) where is the agent in a world whose geometry is known? We call these problems, '*where is it?*', '*what is it?*', and '*where am I?*'.

The Iterative Closest Point (ICP) algorithm is often used to address these problems. It was introduced by [1] for point cloud matching and is a frequently applied method in object pose estimation [2,3], object classification [4,5] and localisation [6,7]. The ICP method minimises the error between a point cloud and a geometric model [8]. The motivation for minimising error (or other error-based metrics) is that an object, when placed in the correct pose, should reproduce the observed range or point cloud measurements (i.e., zero error). A limitation of ICP is that it is not robust if the point cloud is unsegmented or noise is present.

The method of this paper, which we call Maximum Sum of Evidence, overcomes this limitation by seeking to determine the pose that is most evidenced in the point cloud. The paper builds from [9,10] and demonstrates the efficacy of an evidenced-based approach to addressing '*where is it?*', '*what is it?*', and '*where am I?*' problems.

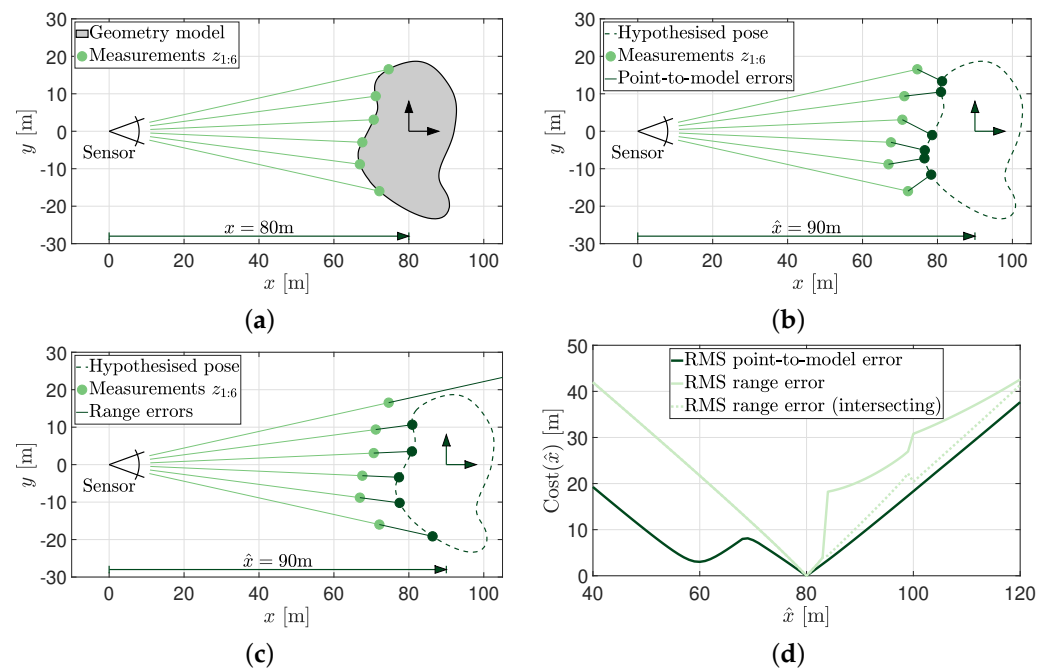
The paper is organised as follows. Section 2 articulates the limitations of using error-based minimisation. Section 3 details the MSOE algorithm in two parts: (i) Section 3.1 describes an objective function based on the conditional likelihood of range measurements; (ii) Section 3.2 explores searching heuristics that employ the described objective function. The algorithm is demonstrated in Section 4 for its ability to answer '*where is it?*', '*what is it?*', and '*where am I?*' type questions that frequently arise as required perception capabilities

in robotics and automation. Section 5 further evaluates the performance of the MSOE estimator in examples that are known to challenge ICP. Concluding remarks are provided in Section 6.

## 2. The Limitations of Using Error-Based Minimisation in Pose Estimation Problems

Error-based minimisation is commonly employed in pose estimation problems where point cloud or range measurements are provided as input [11,12]. This is intuitive as an estimate that reproduces the observed range measurements,  $Z$ , will result in zero cost. As such, a search heuristic can be applied that is guided by an objective function that seeks to find a pose hypothesis that minimises an error-based metric. However, the results are prone to error in real data sets.

Consider, for example, the object of known geometry shown in Figure 1a which is free to move in the  $x$ -coordinate towards or away from the sensor. The geometry, in truth, is located at  $x = 80$  m which produces the range measurements,  $z_{1:6}$ . Two cost-based metrics could readily be applied to this problem to estimate the lateral position of the geometry. The first is illustrated in Figure 1b whereby the point-to-model cost of a measurement is calculated as the Euclidean distance to the closest surface on the geometry model located by the position hypothesis,  $\hat{x}$ . The overall cost of the hypothesis,  $\text{cost}(\hat{x})$ , can be taken as the sum or root-mean-square (RMS) of these errors. The point-to-model errors are shown for a hypothesis of  $\hat{x} = 90$  m. Figure 1c illustrates the use of range measurement error for the same position hypothesis. Here, the cost is measured as the difference between actual observations and expected observations. Note that the first range measurement,  $z_1$ , no longer hits the geometry in the  $\hat{x} = 90$  m position, and has its error measured against the maximum range of the sensor.



**Figure 1.** Cost-based metrics and their ability to provide for the estimation of object pose when provided with a perfect measurement set. (a) Six range measurements of a geometric body located at  $x = 80$  m. (b) Point-to-model errors when the geometry model is hypothesised to be located at  $\hat{x} = 90$  m. (c) Range measurement errors when the geometry model is hypothesised to be located at  $\hat{x} = 90$  m. (d) RMS point-to-model and range error cost metrics for hypotheses  $[40 \text{ m} \leq \hat{x} \leq 120 \text{ m}]$ .

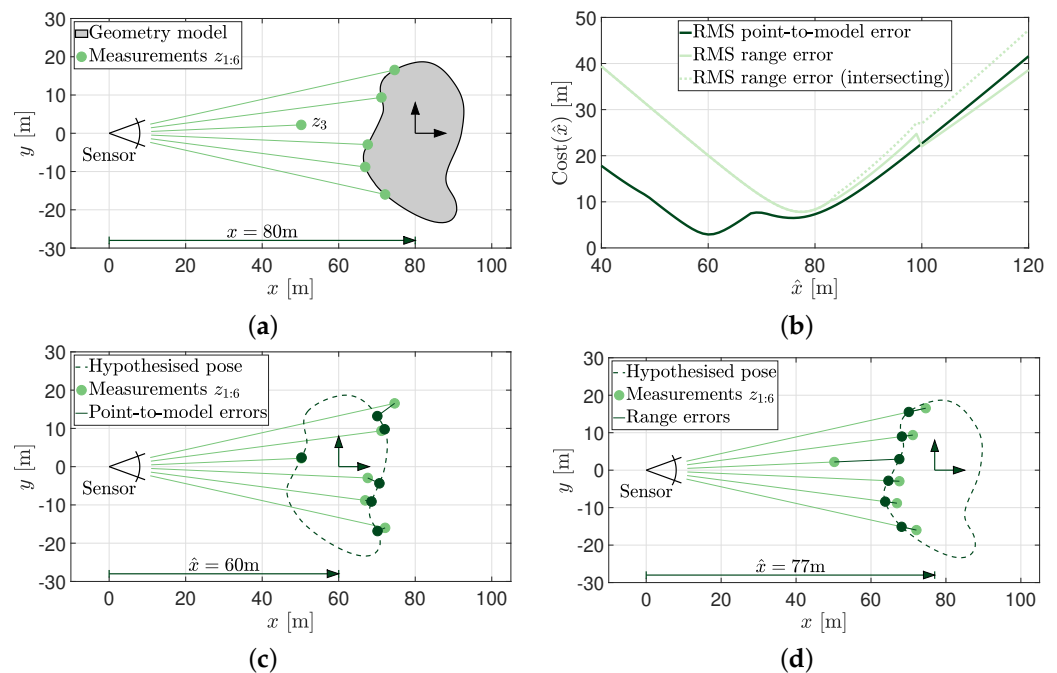
Equal arguments can be made for either removing or retaining the non-intersecting beam. Retaining the measurement will cause a discontinuity in the cost function as the RMS of range error will rise significantly when beams ‘fall off’ the model. Removing the

non-intersecting measurement will misleadingly reduce the cost as a hypothesis with no intersecting beams would have theoretically no cost.

Figure 1d shows the cost of position hypotheses between 40 m to 120 m. The dark and light lines indicate the RMS of point-to-model and range errors, respectively. A dashed light line has also been included to indicate the RMS of range errors when only intersecting beams are considered. This figure highlights the immediate attractiveness of using cost-metrics. The globally minimum cost of both cost-metrics is associated with the correct pose of  $x = 80$  m. Both cost metrics, as expected, compute the cost of the true pose to be zero.

At surface level, cost-based metrics are appealing as they appear to provide for the correct pose solution. This paper, however, argues that cost-based metrics are an inherently poor metric for measuring hypothesis correctness. This is most easily demonstrated when measurement uncertainty is introduced to the same example problem.

Figure 2 shows the same pose estimation problem demonstrated in Figure 1, however, a single measurement,  $z_3$ , has been corrupted (Figure 2a). Despite having five perfect measurements of the geometry, this one incorrect measurement is enough to significantly affect the overall cost and change the estimated pose solution. The objective functions when applied to this slightly altered measurement set have been plotted in Figure 2b. Note that there is no longer a zero cost solution as no position could possibly recreate the complete set of observed measurements. More importantly, the globally minimum costs are no longer located at the true solution. The point-to-model error cost metric suggests that the best location for the geometry is at  $\hat{x} = 60$  m (shown in Figure 2c). Likewise, the range error cost metric suggests that a pose of  $\hat{x} = 77$  m will result in a globally minimum cost (shown in Figure 2d).



**Figure 2.** Cost-based metrics and their ability to provide for the estimation of object pose when provided with a corrupted measurement set. (a) One of the range measurements,  $z_3$ , is altered to simulate a spurious measurement. (b) RMS point-to-model and range error cost metrics when the measurement  $z_3$  is altered. (c) The most likely pose hypothesis using the point-to-model cost metric. (d) The most likely pose hypothesis using the range error cost metric.

This example demonstrates that cost-based metrics are prone to fail because they are designed to determine where an object must lie in order to recreate the complete measurement set. In real data the assumption that all measurements must belong to the object can not be justified.

Rather than ask ‘Where does the geometry need lie to recreate the measurements?’ a better question is ‘What pose do these measurements support or provide the most evidence towards?’. This is the fundamental difference in ideology that is adopted by this paper.

### 3. The Maximum Sum of Evidence (MSOE) Method

The Maximum Sum Of Evidence method is predicated on the belief that the most likely pose hypothesis,  $\mathbf{H}^*$ , is that which is most supported by the individual range measurement observations. Each of the  $N$  observed range measurements,  $\mathbf{Z} = \{z_1, z_2, \dots, z_N\}$ , will support a region of the pose hypothesis space,  $\mathbf{H}$ . We consider a discretised hypothesis space of  $M$  hypotheses,  $\mathbf{H} = \{\mathbf{H}_1, \mathbf{H}_2, \dots, \mathbf{H}_M\}$ .

The law of total probability relates the marginal probability of the  $j$ -th pose hypothesis,  $P(\mathbf{H}_j)$ , to its conditional probability given each range observation,  $z_i$ . It is expressed as,

$$P(\mathbf{H}_j) = \sum_{i=1}^N P(\mathbf{H}_j|z_i)P(z_i). \quad (1)$$

The MSOE method seeks to determine the most probable hypothesis,  $\mathbf{H}^*$ , i.e., that which maximises the total probability measure. The most evidenced hypothesis is indexed by  $j^*$  (i.e.,  $\mathbf{H}^* = \mathbf{H}_{j^*}$ ), which is determined as,

$$j^* = \operatorname{argmax}_j \{P(\mathbf{H}_j)\} = \operatorname{argmax}_j \left\{ \sum_{i=1}^N P(\mathbf{H}_j|z_i)P(z_i) \right\}. \quad (2)$$

The conditional probability of a hypothesis given a range observation,  $P(\mathbf{H}_j|z_i)$ , is related to the conditional probability of the hypothesis providing the range observation,  $P(z_i|\mathbf{H}_j)$  via Bayes’ theorem. It can be expressed as,

$$P(\mathbf{H}_j|z_i) = \frac{P(z_i|\mathbf{H}_j)P(\mathbf{H}_j)}{P(z_i)}. \quad (3)$$

Substituting Bayes’ theorem (Equation (3)) into the expression for total probability (Equation (2)) yields,

$$j^* = \operatorname{argmax}_j \left\{ \sum_{i=1}^N \frac{P(z_i|\mathbf{H}_j)P(\mathbf{H}_j)}{P(z_i)} P(z_i) \right\} = \operatorname{argmax}_j \left\{ \sum_{i=1}^N P(z_i|\mathbf{H}_j)P(\mathbf{H}_j) \right\}. \quad (4)$$

There are two simplifications that can be made to this equation. The first, is that, in the absence of prior information, all hypotheses can be considered equally likely, i.e.,

$$P(\mathbf{H}_j) = 1/M. \quad (5)$$

The second simplification comes from recognising that range measurements have a finite resolution,  $\Delta z$ , and are therefore discrete. The discrete probability of a range measurement can be expressed as,

$$P(z_i|\mathbf{H}_j) = \int_{z=z_i-\Delta z/2}^{z=z_i+\Delta z/2} f(z_i|\mathbf{H}_j) dz, \quad (6)$$

which, for small  $\Delta z$ , approximates to,

$$P(z_i|\mathbf{H}_j) \simeq f(z_i|\mathbf{H}_j) \cdot \Delta z. \quad (7)$$

Making use of these two simplifications, Equation (4) becomes

$$j^* = \operatorname{argmax}_j \left\{ \frac{\Delta z}{M} \sum_{i=1}^N f(z_i|\mathbf{H}_j) \right\}. \quad (8)$$

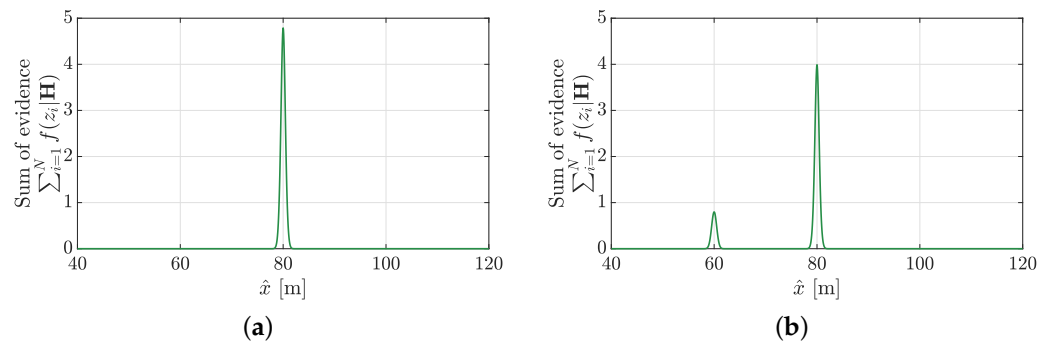


The scalar terms  $M$  and  $\Delta z$  do not relocate the maximum and can be removed, leaving,

$$j^* = \underset{j}{\operatorname{argmax}} \left\{ \sum_{i=1}^N f(z_i | \mathbf{H}_j) \right\}. \quad (9)$$

This result forms the basis of the MSOE algorithm. The most probable hypothesis,  $\mathbf{H}^*$ , is that which maximises the sum of conditional measurement likelihoods,  $f(z_i | \mathbf{H}_j)$ .

Rather than seek to minimise cost of point cloud error, the MSOE method seeks to maximise the conditional likelihood of obtaining the point cloud observations. This provides a significant advantage towards pose estimation in non-segmented point cloud data (which is demonstrated in Section 5.1). Measurements that are not possible under a hypothesis (i.e.,  $f(z_i | \mathbf{H}_j) = 0$ ) will not add support to the hypothesis but will not detract from its likelihood either. Figure 3 shows the evidence metric applied to the measurement sets from the previous section (Figures 1a and 2a). In the first example, all measurements are shown to support the hypothesis that the object is located at  $\hat{x} = 80$  m. When applied to the second example, the corrupted measurement,  $z_3$ , provides evidence towards the  $\hat{x} = 60$  m hypothesis, but overall the original measurements still support the true solution.



**Figure 3.** The evidence-based metric suggests that the hypothesis  $\hat{x} = 80$  m is most likely even when the spurious measurement is introduced. (a) Evidence measured across the hypothesis space using the perfect measurement set in Figure 1a. (b) Evidence measured across the hypothesis space using the corrupted measurement set in Figure 2a.

The two sections that follow provide detail on how to: (i) calculate the conditional range measurement likelihoods,  $f(z_i | \mathbf{H}_j)$ ; and (ii) guide searches in the hypothesis space  $\mathbf{H}$  for the maximum-sum-of-evidence solution,  $\mathbf{H}^*$ .

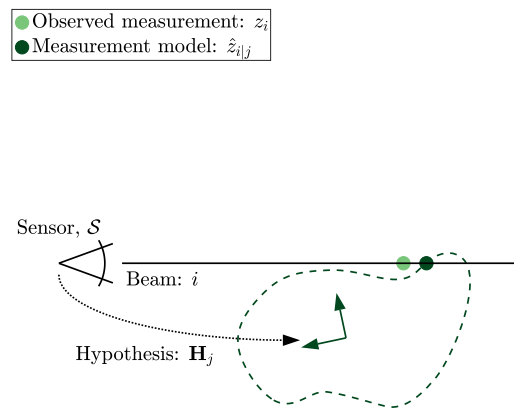
### 3.1. Establishing Conditional Range Measurement Likelihoods, $f(z_i | \mathbf{H}_j)$

The conditional measurement likelihood  $f(z_i | \mathbf{H}_j)$  denotes the conditional likelihood of observing the measurement  $z_i$  on the  $i$ -th beam given an assumed object pose  $\mathbf{H}_j$ . The calculation of this conditional likelihood requires the use of a measurement model to determine the expected range measurement, denoted  $\hat{z}_{i|j}$ . The LiDAR measurement model (denoted  $h(\cdot)$ ) is a function that returns the ray-cast range of a beam against the hypothesis-located geometry model. The  $j$ -th pose hypothesis is modelled by locating the a priori geometry model using  $\mathbf{H}_j$ . The range of the  $i$ -th beam ray-cast against the  $j$ -th pose hypothesis is expressed as

$$\hat{z}_{i|j} = h_i(\mathbf{H}_j). \quad (10)$$

Figure 4 illustrates the ray-cast range measurement  $\hat{z}_{i|j}$  alongside an actual range measurement  $z_i$ . Here, the hypothesis,  $\mathbf{H}_j$ , locates the geometry relative to the sensor coordinate frame,  $\mathcal{S}$ .

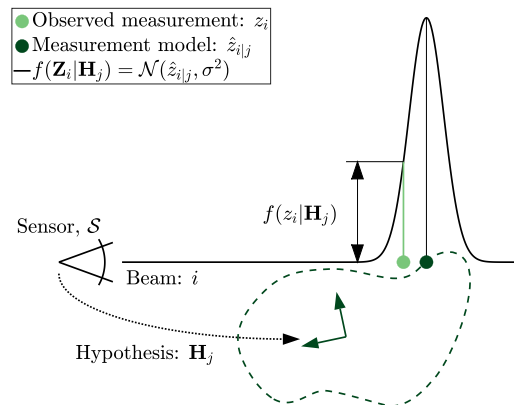
The support that the observed measurement,  $z_i$ , provides to the hypothesis is determined through the use of a conditional range likelihood model,  $f(z_i | \mathbf{H}_j)$ . The conditional range likelihood model describes where we might expect measurements to occur if the hypothesis were true.



**Figure 4.** The expected range along the  $i$ -th beam for hypothesis  $H_j$  is depicted as  $\hat{z}_{ij}$ .

A starting point is to consider the measurement uncertainty of the LiDAR sensor itself. Manufacturers will typically provide the standard deviation of sensor measurement uncertainty,  $\sigma$ .

Consider Figure 5 which illustrates the range measurement distribution considering only the sensor's measurement uncertainty. The measurement uncertainty has been exaggerated to  $\sigma = 3$  m for illustrative purposes. In this example, the observed range measurement along this beam,  $z_i$ , is 60 m and is indicated to be a slightly shorter than the range,  $\hat{z}_{ij} = 64$  m.



**Figure 5.** Range likelihood distribution,  $f(\mathbf{Z}_i|\mathbf{H}_j)$ , given sensor measurement uncertainty,  $\sigma$ .

The likelihood of observing the actual measurement,  $z_i$ , given the pose hypothesis,  $H_j$ , and the sensor's measurement uncertainty,  $\sigma$ , is calculated as,

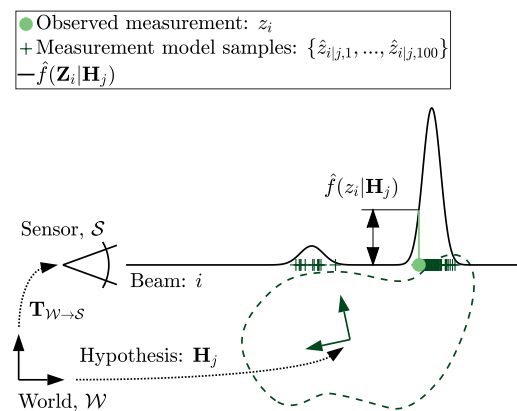
$$f(z_i|\mathbf{H}_j) = \mathcal{N}(\hat{z}_{ij}, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left(-\frac{(z_i - \hat{z}_{ij})^2}{2\sigma^2}\right). \quad (11)$$

In this example, a range measurement of 64 m would have provided the most evidence towards hypothesis  $H_j$ , however, the observed range of  $z_i = 60$  m still reasonably evidences the hypothesis.

It is possible to map other sources of uncertainty into the range measurement likelihood function,  $f(z_i|\mathbf{H}_j)$ . By way of example, consider the more general problem of estimating the pose of an object within a greater World coordinate system,  $\mathcal{W}$ , in which the pose of the sensor,  $T_{\mathcal{W} \rightarrow S}$ , is also uncertain.

The uncertainty in the sensor's pose relates to how accurate the extrinsic calibration or registration procedure is. In previous studies we have estimated that this can be determined within several millimetres/milliradians for translational and rotational components,

respectively, [13,14]. Figure 6 illustrates the effect of registration uncertainty,  $\text{Cov}(\mathbf{T}_{\mathcal{W} \rightarrow \mathcal{S}})$ , on the range measurement distribution,  $f(\mathbf{Z}_i | \mathbf{H}_j)$ .



**Figure 6.** Range likelihood distribution,  $\hat{f}(\mathbf{Z}_i | \mathbf{H}_j)$ , given sensor measurement uncertainty,  $\sigma$  and sensor pose uncertainty,  $\text{Cov}(\mathbf{T}_{\mathcal{W} \rightarrow \mathcal{S}})$ .

Sensor pose uncertainty will redirect the beam onto other parts of the geometry model as located by the hypothesis. The redirection of beams can result in abrupt changes to the expected range measurement. The resulting distribution of range measurements now depends on the shape of the geometry. An estimate of the range probability density function,  $\hat{f}(\mathbf{Z}_i | \mathbf{H}_j)$ , can be established by sampling the range measurement model under perturbed sensor poses drawn from extrinsic calibration uncertainty,  $w$ , and adding the sensor's measurement uncertainty,  $v$ , to the result. The sampled range measurements along the  $i$ -th beam under the  $j$ -th pose hypothesis,  $\hat{\mathbf{Z}}_{i|j} = \{\hat{z}_{i|j,1}, \hat{z}_{i|j,2}, \dots, \hat{z}_{i|j,n}\}$ , are each calculated from

$$\hat{z}_{i|j,k} = h_i(\mathbf{H}_j, \mathbf{w}_k) + v_k, \quad (12)$$

where  $\mathbf{w}_k \sim \mathcal{N}(\mathbf{0}, \text{Cov}(\mathbf{T}_{\mathcal{W} \rightarrow \mathcal{S}}))$  and  $v_k \sim \mathcal{N}(0, \sigma^2)$ .

Figure 6 shows the placement of  $n = 100$  sampled range measurements under this new measurement model. The distribution is shown to contain a new mode where the sensor's pose uncertainty,  $\text{Cov}(\mathbf{T}_{\mathcal{W} \rightarrow \mathcal{S}})$ , could redirect the beam onto a different part of the object's geometry.

Much work has been done towards the estimation of probability density functions (PDFs) from sampled data, see for example [15]. One such technique is to construct a Gaussian kernel density estimator (KDE) where a small kernel,  $\mathbf{K}(\cdot)$ , is located at each of the ray-cast samples,  $\hat{z}_{i|j,k}$ . The estimated range probability density function,  $\hat{f}(\mathbf{Z}_i | \mathbf{H}_j)$ , evaluated at  $z_i$  is then calculated from the summation of kernels

$$\hat{f}(z_i | \mathbf{H}_j) = \frac{1}{n \cdot h} \sum_{k=1}^n \mathbf{K}\left(\frac{z_i - \hat{z}_{i|j,k}}{h}\right), \quad (13)$$

where the Gaussian kernels are described by

$$\mathbf{K}(x) = \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{x^2}{2}\right). \quad (14)$$

The parameter  $h$  is commonly referred to as the *bandwidth*, and acts as a smoothing parameter between bias and variance. The optimal selection of  $h$  greatly influences the accuracy to which  $\hat{f}(\cdot)$  represents the true density  $f(\cdot)$ . Turlach provides a good review of optimal bandwidth techniques in [16]. The performance of  $\hat{f}(\cdot)$  to a selected bandwidth,  $h$ , can be measured by the integrated squared error (ISE) over the estimated function's domain

$$\text{ISE}(h) = \int \left(\hat{f}(\mathbf{Z}_i; h | \mathbf{H}_j) - f(\mathbf{Z}_i | \mathbf{H}_j)\right)^2 d\mathbf{Z}_i. \quad (15)$$

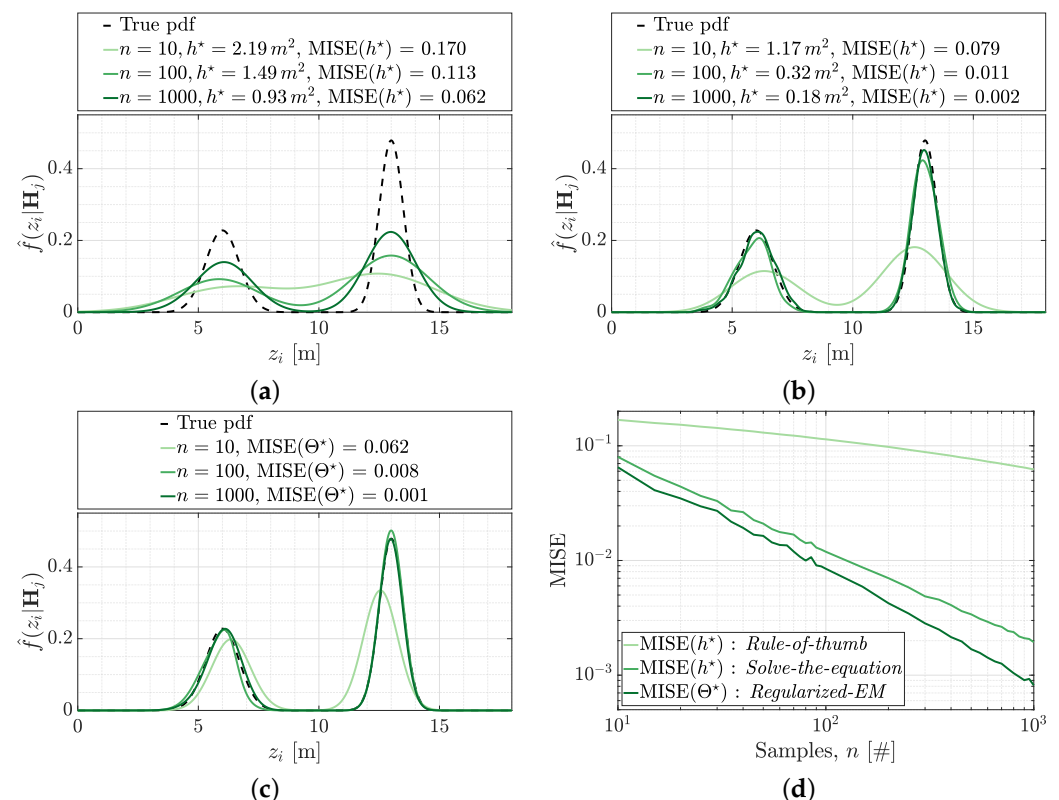
However, this criterion is subject to the sampled range measurement predictions,  $\hat{Z}_{ij}$ , which are themselves subject to the sampled perturbations,  $w_k$  and  $v_k$ , used during the raycasting process described in Equation (12). A true assessment of the density estimate's accuracy is obtained from the expected value of the ISE or *mean integrated squared error* (MISE)

$$\text{MISE}(h) = \mathbb{E} \left[ \int \left( \hat{f}(\mathbf{Z}_i; h | \mathbf{H}_j) - f(\mathbf{Z}_i | \mathbf{H}_j) \right)^2 d\mathbf{Z}_i \right]. \quad (16)$$

For Gaussianly distributed sample range measurements,  $\hat{Z}_{ij}$ , the optimal bandwidth (for minimising  $\text{MISE}(h)$ ) can be found from Silverman's plugin *rule-of-thumb* [17] equation,

$$h^* = \left( \frac{4\hat{\sigma}^5}{3n} \right)^{1/5}, \quad (17)$$

where  $\hat{\sigma}$  is the standard deviation of sampled range measurements,  $\hat{Z}_{ij} = \{\hat{z}_{ij,1}, \dots, \hat{z}_{ij,n}\}$ . It is not reasonable to expect the range measurement likelihood  $f(\mathbf{Z}_i | \mathbf{H}_j)$  to be Gaussian, and this bandwidth selection is known to perform poorly on multimodal PDFs (which are shown to arise in Figure 6). Figure 7a demonstrates the performance of kernel density estimation using this optimal bandwidth,  $h^*$ , towards the estimation of a bimodal probability density function  $f(\mathbf{Z}_i | \mathbf{H}_j) = 0.4 \cdot \mathcal{N}(6, 0.7^2) + 0.6 \cdot \mathcal{N}(13, 0.5^2)$ . The estimated PDF,  $\hat{f}(\mathbf{Z}_i; h^* | \mathbf{H}_j)$ , is shown to be oversmoothed, however, the MISE does decrease as the number of samples  $n$  increases.



**Figure 7.** The mean integrated squared error of the estimated conditional range measurement likelihood function,  $\hat{f}(\mathbf{Z} | \mathbf{H}_j)$ , as determined using: (a) Kernel density estimation using optimal bandwidth,  $h^*$ , established from *Rule-of-thumb* [17]. (b) Kernel density estimation using optimal bandwidth,  $h^*$ , established from *Solve-the-equation* [18]. (c) Density estimation using optimal parameters,  $\Theta$ , established from regularised expectation maximisation [19]. (d) The MISE for the three estimates are shown for increasing number of samples,  $n$ .

Sheather and Jones present an improved bandwidth selection method that is guided by the sample data [20]. The optimal bandwidth solution,  $h^*$ , seeks to minimise good quality estimates of the MISE. The method is commonly referred to as a *solve-the-equation* approach. Botev et al. improve on this method by removing the assumption that the underlying data is parametric [18]. Removing this assumption prevents the estimator from deteriorating for multimodal densities with widely separated modes. Figure 7b shows the estimated  $\hat{f}(\mathbf{Z}_i; h^* | \mathbf{H}_j)$  using the improved *solve-the-equation* approach of [18].

Expectation maximisation, or EM, provides an alternative way to estimate the probability density function,  $f(\mathbf{Z}_i | \mathbf{H}_j)$ . Rather than locating kernels on sample data, EM is an iterative method that seeks to recover the parameters,  $\Theta$ , describing the underlying modes/species of the sample data,  $\hat{\mathbf{Z}}_{ij} = \{\hat{z}_{ij,1}, \dots, \hat{z}_{ij,n}\}$ . This is found by maximising the log-likelihood function describing the likelihood of obtaining each range measurement sample,  $\hat{z}_{ij,k}$ , given the assumed parameter model,  $\Theta$ ,

$$\mathcal{L}(\Theta; \hat{\mathbf{Z}}_{ij}) = \sum_{k=1}^n \log(f(\hat{z}_{ij,k} | \Theta)). \quad (18)$$

The example PDF used in Figure 7 is a Gaussian mixture model constructed from two modes. The first mode provides 40% ( $\alpha_1$ ) of the density and is parameterised with  $\mu_1 = 6$  m and  $\sigma_1 = 0.7$  m<sup>2</sup>, the second mode provides the remaining 60% ( $\alpha_2$ ) and is parameterised by  $\mu_2 = 13$  m,  $\sigma_2 = 0.5$  m<sup>2</sup>. The parameter space to be estimated is  $\Theta = \{\alpha_1, \mu_1, \sigma_1, \alpha_2, \mu_2, \sigma_2\}$ .

A common criticism of EM estimation methods is that the solution must assume a number of underlying modes in the density being estimated and this is usually referred to as the model selection problem. An EM implementation that sought to model the example PDF with three modes would not be able to reduce the over specified model back to two modes. Li et al. [19] provide an EM implementation that can reduce over-specified models by effectively reducing component weightings,  $\alpha$ , to zero. The implementation is called the *Regularised EM* algorithm as the likelihood measure (Equation (18)) is regularised by applying a penalty based on the mutual information between sampled *observation data*,  $\hat{\mathbf{Z}}_{ij}$ , and the corresponding *missing data*,  $\mathcal{Y}$ . Figure 7c shows the estimated PDF using regularised expectation maximisation.

The conditional range measurement likelihood can be estimated using kernel density estimation,  $\hat{f}(z_i; h^* | \mathbf{H}_j)$ , or regularised expectation maximisation,  $\hat{f}(z_i; \Theta^* | \mathbf{H}_j)$ , with both showing low MISE values for reasonably small sample sizes,  $n$ . Figure 7d shows that the kernel density estimate typically requires 1.5–2.5 as many samples as the regularised expectation maximisation method to achieve the same MISE.

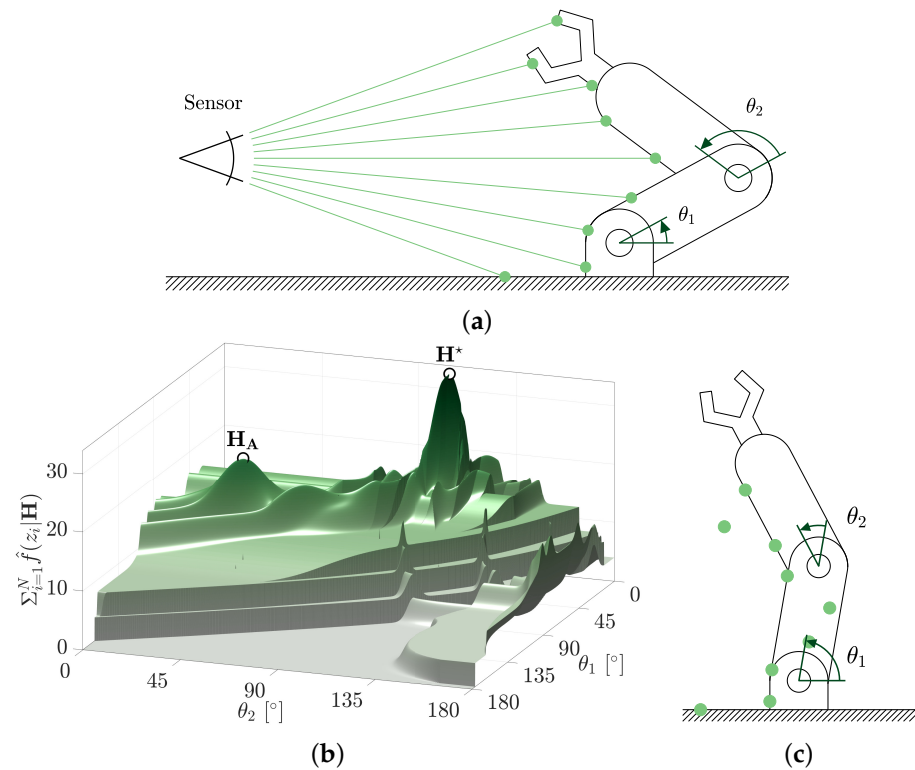
### 3.2. Searching the Hypothesis Space for the Global Maximum Sum of Evidence, $\mathbf{H}^*$

The sum of evidence measure described in Equation (9) requires the summation of conditional measurement likelihoods. As this paper will show, the evidence metric serves as a very robust objective function for the measure of hypothesis likelihood. A search heuristic is required to determine the hypothesis parameters that yield the maximum sum of evidence. This section explores search heuristic performance using an example 2D search space of a robotic manipulator.

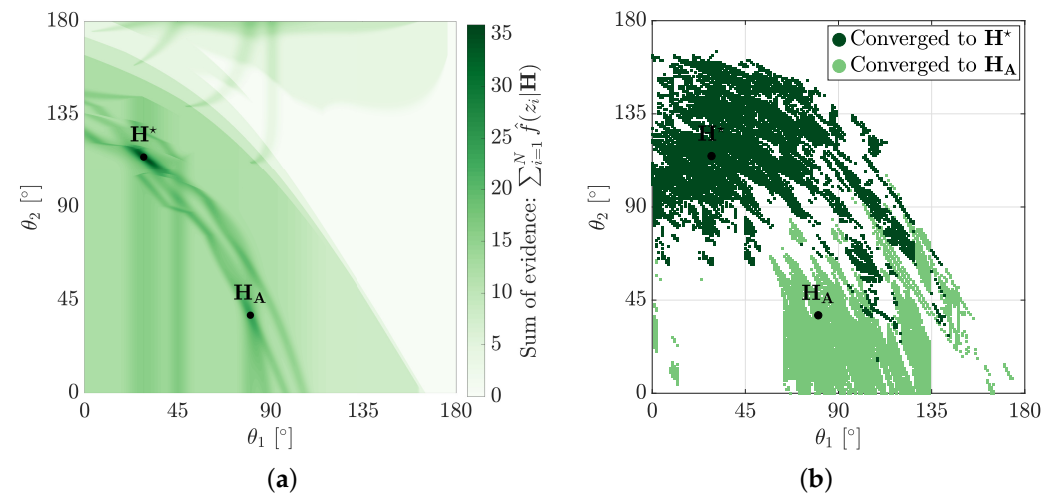
The example robot geometry is parameterised by the hypothesis space  $\mathbf{H} = [0^\circ \leq \theta_1 \leq 180^\circ, 0^\circ \leq \theta_2 \leq 180^\circ]$  (Figure 8a). Nine range measurements,  $\mathbf{Z}$ , are obtained from a sensor with the robot located by  $\mathbf{H}_{truth} = [28.6^\circ, 114.6^\circ]$ . Figure 8b shows the evidence measure of Equation (9) mapped across the complete hypothesis space,  $\mathbf{H}$ . The true hypothesis is shown to produce the global maximum sum of evidence,  $\mathbf{H}^*$ , at the correct location,  $\mathbf{H}_{truth}$ . An immediate concern, however, is that many local maximums of evidence exist within this surface. Other robot configurations, such as  $\mathbf{H}_A$  (Figure 8c), are reasonably well evidenced from the range measurements and will result in local maximums.

Incorrect pose hypotheses (such as  $\mathbf{H}_A$ ) can have considerably large basins of convergence. Figure 9b shows the basin of convergence for hypotheses  $\mathbf{H}^*$  and  $\mathbf{H}_A$ . The basins

were determined by examining the solutions of a Nelder-Mead solver (with side length  $10^\circ$ ) seeded in  $1^\circ$  increments of the hypothesis space,  $\mathbf{H}$ . Approximately 18% of the seed locations are shown to converge to the MSOE hypothesis,  $\mathbf{H}^*$ , however, 11% were also shown to converge to the competing local maximum,  $\mathbf{H}_A$ . The remaining 71% of seeds converged elsewhere or not at all.



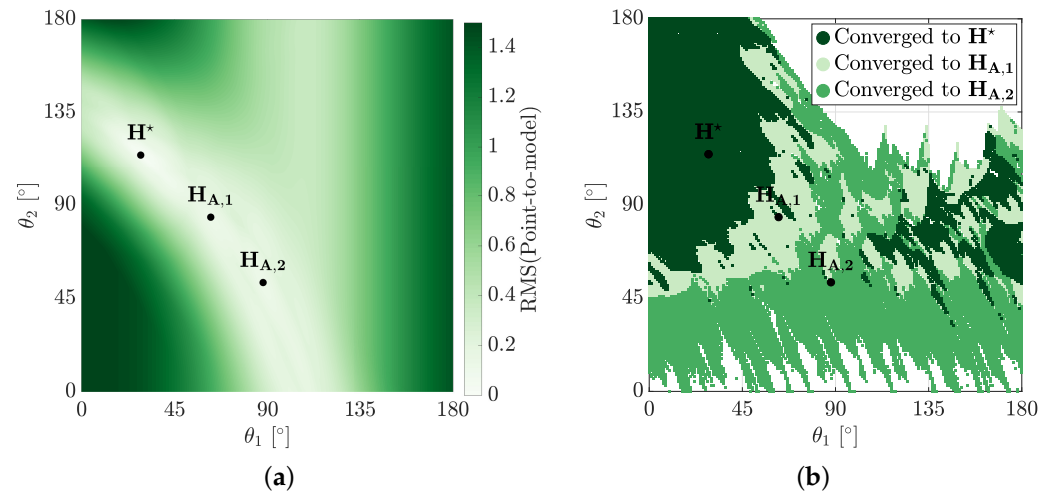
**Figure 8.** An example 2-DOF problem used to explore the challenges of searching the hypothesis space,  $\mathbf{H}$ , for the global maximum sum of evidence solution,  $\mathbf{H}^*$ . (a) Geometry and measurements of the example two degree of freedom estimation problem. (b) The sum of conditional likelihoods of the nine range measurements evaluated across the hypothesis space,  $\mathbf{H} = [\theta_1, \theta_2]$ . (c) Incorrect local maximum sum of evidence at pose  $\mathbf{H}_A$ .



**Figure 9.** The convergence basin for the MSOE pose,  $\mathbf{H}^*$ , and a competing local maximum,  $\mathbf{H}_A$ . The basins are determined using a Nelder-Mead solver seeded at every  $1^\circ$  of  $\theta_1$  and  $\theta_2$ . (a) Evidence measured over the hypothesis space. (b) Nelder-Mead seed convergence using the evidence-based metric.



The local maximums that appear in the evidence-based metric are not dissimilar to the local minimums that arise in cost-based metrics. Figure 10a shows the RMS of point-to-model costs over the same hypothesis space.  $H^*$  is shown to correctly occur at the global minimum, however, two competing hypotheses  $H_{A,1}$  and  $H_{A,2}$  are shown to produce relatively large basins of convergence (Figure 10b).



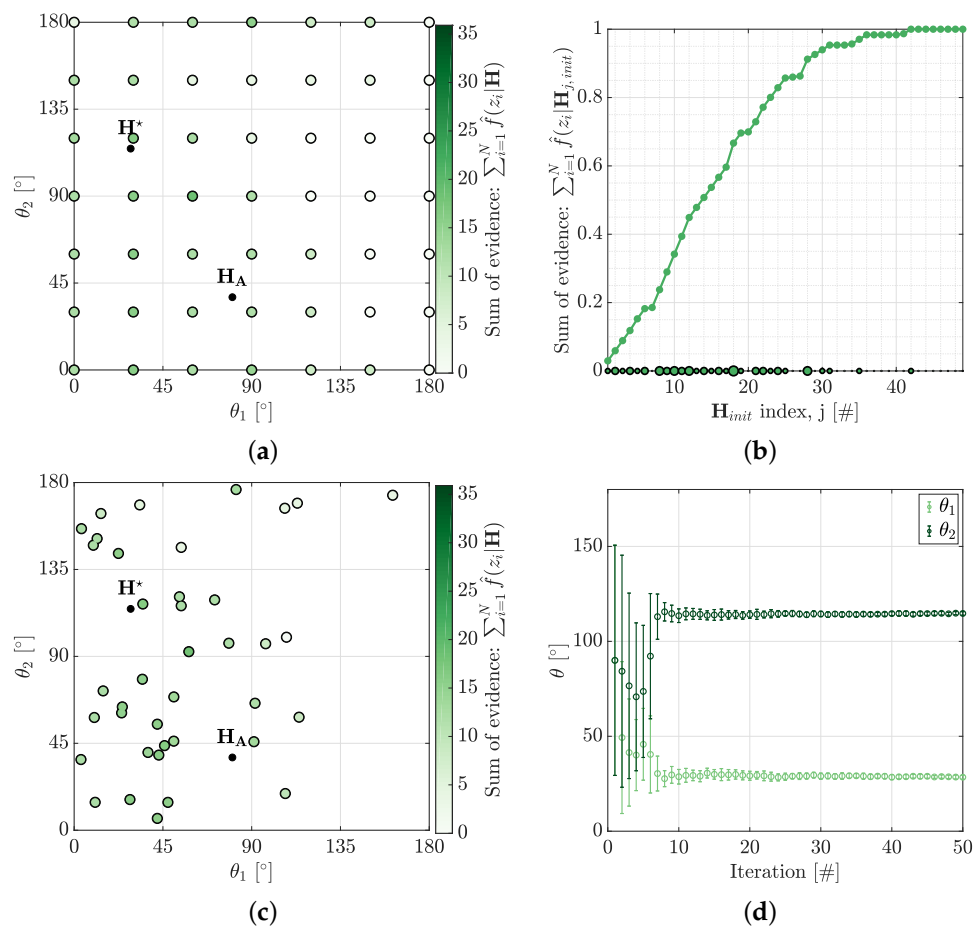
**Figure 10.** The RMS of point-to-model cost metric has a global minimum at,  $H^*$ , and two competing local minimums at,  $H_{A,1}$  and  $H_{A,2}$ . The basins are determined using a Nelder-Mead solver seeded at every  $1^\circ$  of  $\theta_1$  and  $\theta_2$ . (a) RMS point-to-model error. (b) Nelder-Mead seed convergence on the RMS of point-to-model cost metric.

The large convergence basins for locally optimal hypotheses promote the need for an alternate searching heuristic that is not seed-dependent. An effective approach is to use a particle-based heuristic to iteratively search the hypothesis space for the MSOE hypothesis,  $H^*$ . The hypothesis space is uniformly sampled over regular intervals,  $\Delta H$ , to establish an initial hypothesis set,  $H_{init}$ . Figure 11a shows an initial 49 hypotheses established using  $\Delta H = [30^\circ, 30^\circ]$ .

The next hypothesis set,  $H_{next}$ , is determined from a cumulative distribution of the evidences determined under  $H_{init}$ . Uniformly sampling the CDF along its  $y$ -axis provides a base,  $H_{base}$ , for where  $H_{next}$  should concentrate resampling (Figure 11b). Higher evidenced hypotheses of  $H_{init}$  will provide steeper gradient in the CDF which will in turn be resampled more. The circle sizes on the  $x$ -axis of Figure 11b indicate the number of times that hypothesis is entered into  $H_{base}$ . The  $j$ -th hypothesis of the next iteration is determined by adding a small perturbation to  $H_{base}$ ,

$$H_{j,next} \sim \mathcal{N}(H_{j,base}, \text{Cov}(\Delta H / \text{iteration})). \quad (19)$$

The covariance,  $\Delta H / \text{iteration}$ , decreases with each iteration, allowing the solution to converge (Figure 11d).



**Figure 11.** Convergence of the particle search method used to locate  $\mathbf{H}^*$ . (a) The sum of evidence is determined for an initial set of hypotheses,  $\mathbf{H}_{init}$ . (b) A CDF is constructed from this initial measure to guide the concentration of the search for (c) the next set of hypotheses,  $\mathbf{H}_{next}$ . (d) The solution for this example is shown to converge to  $\mathbf{H}^*$  in approximately 10 iterations of this process, error bars indicate  $1\sigma$  of the particle distribution.

### 3.3. Algorithm Summary

A summary of the MSOE algorithm is provided in Algorithm 1. The algorithm can be implemented by: (i) only considering range measurement uncertainty; but can also be extended to (ii) include the affect of other measurement uncertainties via the sampling method described in Section 3.1.

**Algorithm 1:** Maximum Sum of Evidence Algorithm.

---

**Input:** Sensor range measurements,  $\mathbf{Z} = \{z_1, z_2, \dots, z_N\}$ .  
**Input:** Initial hypotheses,  $\mathbf{H}_{init} = \{\mathbf{H}_1, \mathbf{H}_2, \dots, \mathbf{H}_M\}$ , sampled with  $\Delta\mathbf{H}$ .  
**Input:** Range measurement uncertainty,  $\sigma$ , and extrinsic calibration uncertainty,  $\text{Cov}(\mathbf{T}_{\mathcal{W} \rightarrow \mathcal{S}})$ .  
**Input:** Observation model,  $h_i(\cdot)$ , that predicts LiDAR sensor range measurements,  $\hat{z}_{i|j}$ , given a sensor registration,  $\mathbf{T}_{\mathcal{W} \rightarrow \mathcal{S}}$ , and an a priori geometry model of the object.  
**Result:** Maximum sum of evidence pose estimate,  $\mathbf{H}^*$ .

```

1 repeat
2   iteration = 1
3   foreach  $\mathbf{H}_j$  in  $\mathbf{H}$  do
4     Initialise the sum of evidence for  $j$ -th hypothesis:
5      $E_j = 0$ .
6     foreach  $z_i$  in  $\mathbf{Z}$  do
7       Either:
8       (i) Only consider measurement uncertainty,  $\sigma$ :
9         Calculate expected range along  $i$ -th beam using Equation (10):
10         $\hat{z}_{i|j} = h_i(\mathbf{H}_j)$ .
11        Evaluate evidence garnered from measurement  $z_i$  using
12        Equation (11):
13        
$$E_{j|i} = f(z_i|\mathbf{H}_j) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left(-\frac{(z_i - \hat{z}_{i|j})^2}{2\sigma^2}\right)$$

14        (ii) Consider multiple sources of uncertainty via a sampling approach:
15        Sample  $n$  expected ranges along  $i$ -th beam using Equation (12):
16         $\hat{z}_{i|j,k} = h_i(\mathbf{H}_j, \mathbf{w}_k) + v_k$ .
17        Estimate conditional range likelihood  $\hat{f}(\mathbf{Z}_i|\mathbf{H}_j)$  using KDE or EM
18        techniques.
19        Evaluate evidence garnered from measurement  $z_i$ :
20         $E_{j|i} = \hat{f}(z_i|\mathbf{H}_j)$ .
21        Add to total evidence for  $j$ -th hypothesis:
22         $E_j = E_j + E_{j|i}$ .
23        Construct normalised CDF of hypothesis evidences,  $\{E_1, E_2, \dots, E_M\}$ .
24        Uniformly sample along CDF  $y$ -axis to establish  $\mathbf{H}_{base}$ .
25        Construct the next set of test hypotheses using Equation (19):
26         $\mathbf{H}_{j,next} \sim \mathcal{N}(\mathbf{H}_{j,base}, \Delta\mathbf{H}/iteration)$ .
27        Increment the iteration counter.
28         $iteration = iteration + 1$ .
29  until converged;
30 Determine the index of the best hypothesis.
31  $j^* = \underset{j}{\operatorname{argmax}}\{E_j\}$ .
32 return  $\mathbf{H}^* = \mathbf{H}_{j^*}$ .
```

---

**4. Sample Problems**

The following section evaluates the accuracy of MSOE solutions against a set of candidate problems. The problems are designed to test the algorithm's capacity to provide three commonly arising perception capabilities: (i) Pose estimation; (ii) Object classification; and (iii) Localisation. These three areas represent 'where is it?', 'what is it?', and 'where am I?' type questions, respectively, which can all be answered via the MSOE algorithm.

#### 4.1. Pose Estimation: Answering ‘Where Is It?’-Type Questions

‘Where is it?’ questions are a commonly arising question in perception where there is an object whose geometry is known and is expected to appear in the scene. In general, robotic agents need to interact with objects and this requires a belief of where the object is located.

Figure 12 demonstrates the evidence-based metric’s ability to estimate the pose of a known 2D geometry. The MSOE pose hypothesis,  $H^*$ , of the 2D ibis shape is shown against the observed 2D range measurements,  $Z$ . The shape’s pose is described in three degrees of freedom (DOF). It can rotate  $\psi$  about its Z-axis after translating  $x, y$  along its X and Y axes. A measurement set consisting of 21 range measurements was generated from the object at a pose of  $H_{truth} = [\psi, x, y] = [70^\circ, 1.4 \text{ m}, 0.5 \text{ m}]$ .

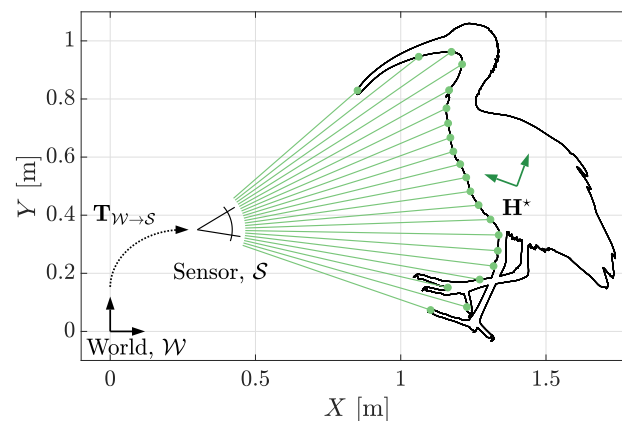


Figure 12. The maximum sum of evidence estimate,  $H^*$ , of a 2D shape’s pose.

The pose solution that generated the maximum sum of evidence was determined to be  $H^* = [70.178^\circ, 1.399 \text{ m}, 0.501 \text{ m}]$ . Comparing  $H_{truth}$  and  $H^*$  directly can be misleading as error in one degree of freedom is often accounted for by error in another. The accuracy of the pose solution is evaluated by determining the maximum geometric displacement of the geometry due to the pose solution error. If points  $P_{truth}$  and  $P^*$  represent the geometry vertices located by  $H_{truth}$  and  $H^*$ , respectively, the maximum geometric error,  $e_{max}$ , is calculated as,

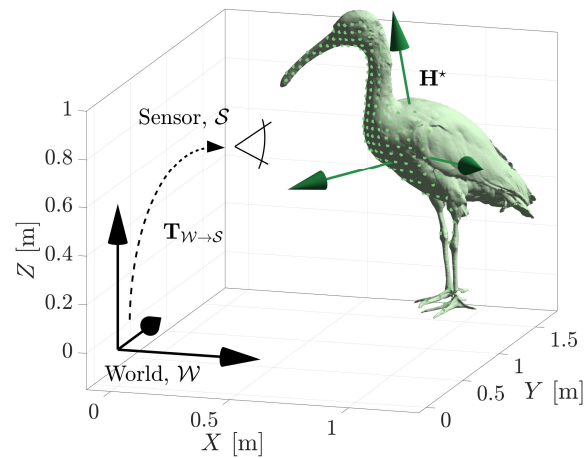
$$e_{max} = \max(|P_{truth} - P^*|). \quad (20)$$

Under the MSOE solution, the maximum geometric displacement of the 2D shape is  $e_{max} = 2.9 \text{ mm}$  (achieved using 10 iterations of the search heuristic described in Section 3.2). At 20 iterations, the maximum displacement is reduced to  $e_{max} = 1.3 \text{ mm}$  and, with the perfect measurement set provided, will, by design, converge to 0 mm.

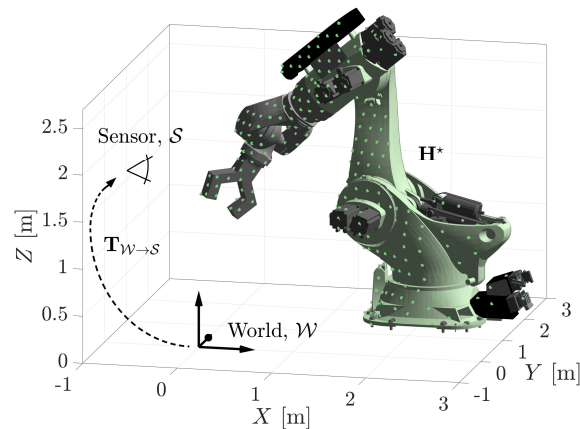
The pose estimation problem can be extended to three-dimensional geometries. Figure 13 shows the MSOE pose solution,  $H^*$ , of a 3D ibis geometry estimated against 3D range data. The object’s pose is described in 6-DOF, it can rotate  $\theta, \phi, \psi$  and translate  $x, y, z$  about and along its X, Y, and Z axes. A measurement set consisting of 150 range measurements was generated from the object at a pose of  $H_{truth} = [\theta, \phi, \psi, x, y, z] = [6^\circ, 8^\circ, 224^\circ, 0.9 \text{ m}, 1.2 \text{ m}, 0.6 \text{ m}]$ . Ten iterations of searching the hypothesis space yielded a MSOE pose solution of  $H^* = [6.48^\circ, 9.95^\circ, 224.23^\circ, 0.900 \text{ m}, 1.203 \text{ m}, 0.605 \text{ m}]$ , which produces a maximum geometric displacement of  $e_{max} = 27.4 \text{ mm}$ . Again, this error will continue to decrease with more iterations.

The region of space occupied by an object does not necessarily have to be parameterised by rigid body transformations. Consider the serial linkage robot shown in Figure 14 which is parameterised by its six rotational degrees of freedom,  $\theta_{1:6}$ . The MSOE algorithm does not require adaptation to joint-space estimation problems. Conditional range measurements likelihoods,  $f(Z|H_j)$ , are still calculated as per Section 3.1. The  $j$ -th hypothesis

$H_j$  is modelled by articulating the individual linkages, rather than applying a rigid frame transformation to the geometry.



**Figure 13.** The maximum sum of evidence estimate,  $H^*$ , of a 3D shape's pose.



**Figure 14.** The maximum sum of evidence estimate,  $H^*$ , of a 6-DOF robot's configuration.

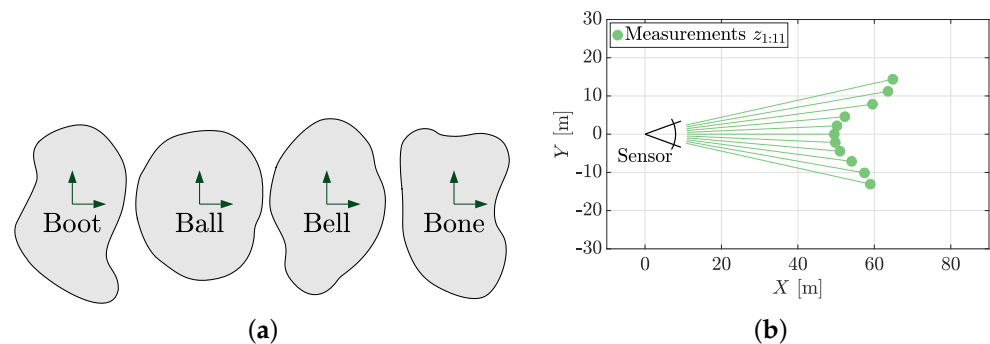
A measurement set of 261 range measurements was generated with the robot configured at  $H_{truth} = [\theta_{1:6}] = [10^\circ, -80^\circ, 120^\circ, -50^\circ, 30^\circ, -10^\circ]$ . After 10 iterations of searching, the estimated solution  $H^* = [10.108^\circ, -79.875^\circ, 119.620^\circ, -52.160^\circ, 31.122^\circ, -2.615^\circ]$  is accurate enough to locate the robot geometry within  $e_{max} = 44.2$  mm of its true position. The joint-space parameters, representing the angular rotations of the serial linkages, are typically increasing in error from  $\theta_1$  to  $\theta_6$ . This is because the estimate of each joint parameter relies on the accuracy of the preceding joints, e.g.,  $\theta_3$  cannot be calculated until a sufficiently accurate  $\theta_1$  and  $\theta_2$  have been obtained. A further 10 iterations of searching reduces the maximum displacement to  $e_{max} = 26.2$  mm.

The MSOE estimator can be applied any parameterisable configuration of geometry that affect range measurements, e.g., object scale.

#### 4.2. Object Classification: Answering 'What Is It?'-Type Questions

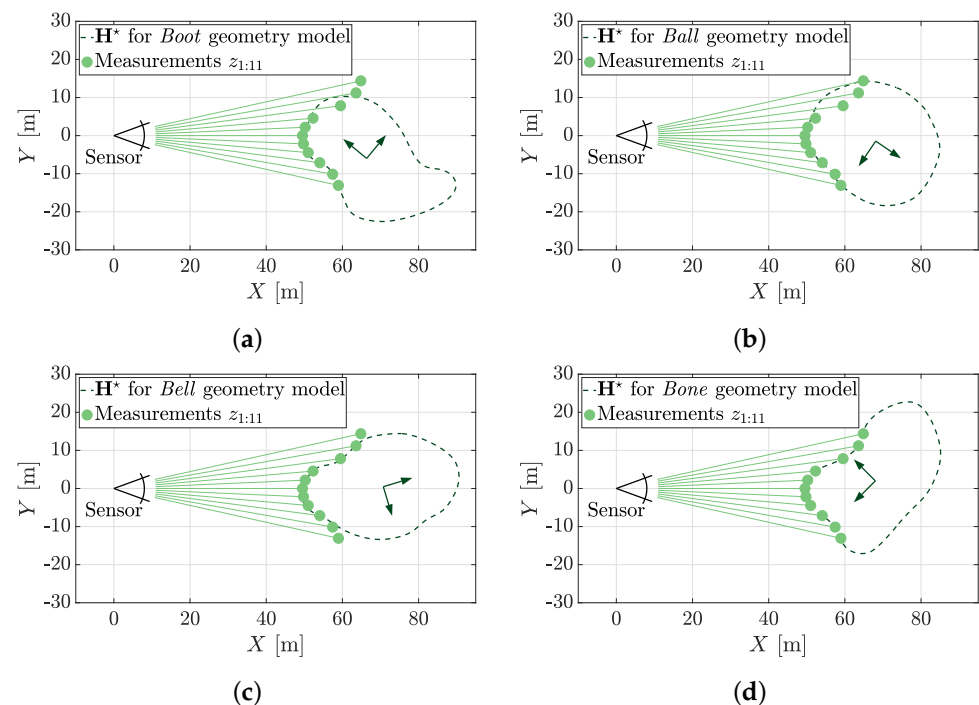
Answering 'What is it?' questions is a commonly required capability that arises in perception-related problems. The sum of conditional measurement likelihoods provided by MSOE is a useful classifier for determining which object is present in a point cloud. The evidence-based metric can be used to identify which shape (from a bank of shape hypotheses) would be least surprising to reproduce a set of observed range measurements. This section demonstrates MSOE for its ability to distinguish between different shapes/geometries in point cloud data.

Figure 15 introduces an example object identification problem in which one of four shapes must be determined from a 2D range measurement set. The *Boot*, *Ball*, *Bell* and *Bone* shapes represent the set of possible shapes for chicken nuggets [21]. The MSOE method can be used to robustly distinguish between them or any other set of geometries where such capacity is needed.



**Figure 15.** This problem seeks to establish which of four nugget geometries is most likely represented by the range measurements  $Z$ . (a) Four chicken nugget shapes. (b) A measurement set,  $Z = \{z_{1:11}\}$ , is produced from the *Bone* nugget.

Figure 16 shows the pose hypotheses,  $H^*$ , that yield the maximum sum of evidence for each nugget geometry model when fitted to the measurement set shown in Figure 15b.

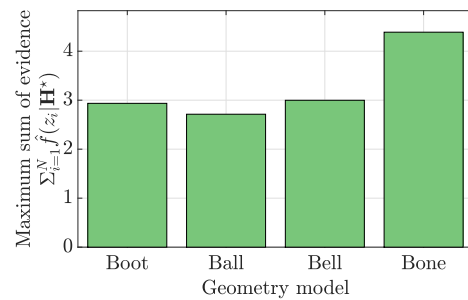


**Figure 16.** Each geometry is fit to the measurement set using the MSOE algorithm. The maximum sum of evidence measured in these four poses is used to determine which geometry is present in the data. (a) The maximum sum of evidence pose of the *Boot* geometry model. (b) The maximum sum of evidence pose of the *Ball* geometry model. (c) The maximum sum of evidence pose of the *Bell* geometry model. (d) The maximum sum of evidence pose of the *Bone* geometry model.

Figure 17 shows the maximum sum of evidence determined for each of the four nugget shapes. The example shows that the *Bone* nugget was most evidenced by the 11 range measurements. In this example, MSOE is demonstrated to simultaneously answer the questions of 'where is it?' and 'what is it?' The discriminating power of the evidence-based

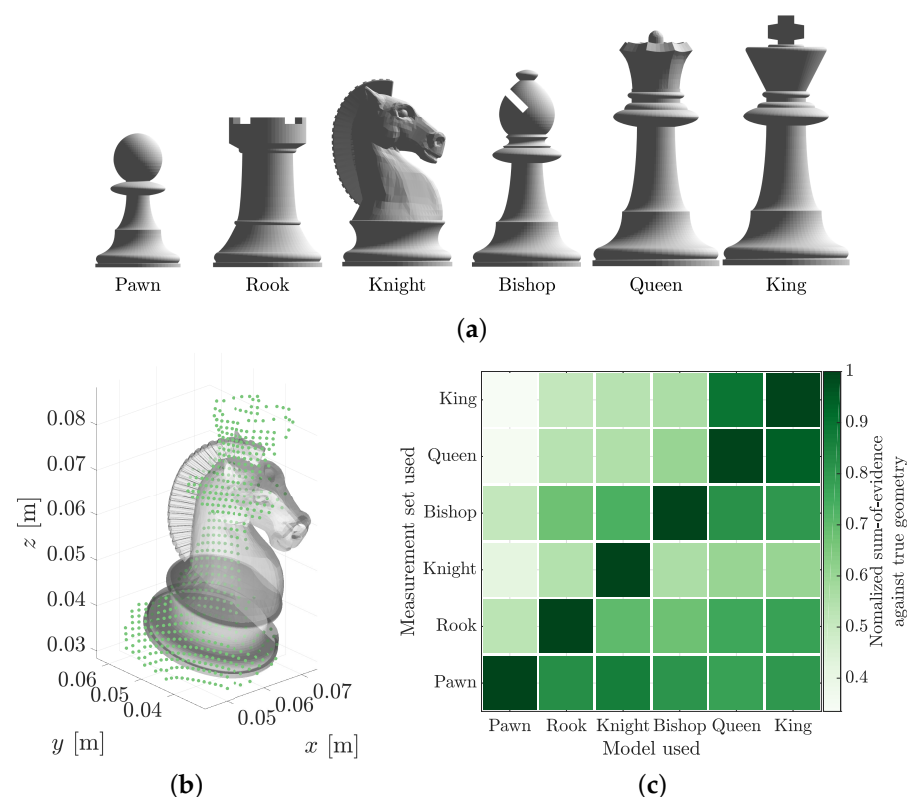


metric towards object identification would be much higher if the pose was already known (i.e., just a ‘what is it?’ question).



**Figure 17.** The maximum sum of evidence for the estimated pose of each geometry model. note that the correct geometry, the *Bone*, is most evidenced by the point cloud data.

The object identification problem can be extended to three-dimensional shapes. Figure 18 explores the problem of classifying 3D chess piece geometries in 3D point cloud data. Six point clouds were established using each geometry model. The MSOE algorithm was used to determine a 6-DOF pose estimate,  $H^*$ , for each geometry model within each of the six point clouds. Figure 18b shows the estimated pose of the *Knight* geometry model within the point cloud established from the *Queen* geometry model. In this case, the MSOE measured with the *Knight* geometry model was only 55% as high as the MSOE measured with the true geometry model, the *Queen*. Figure 18c plots a confusion matrix normalising each sum-of-evidence score against that obtained with the true geometry model.



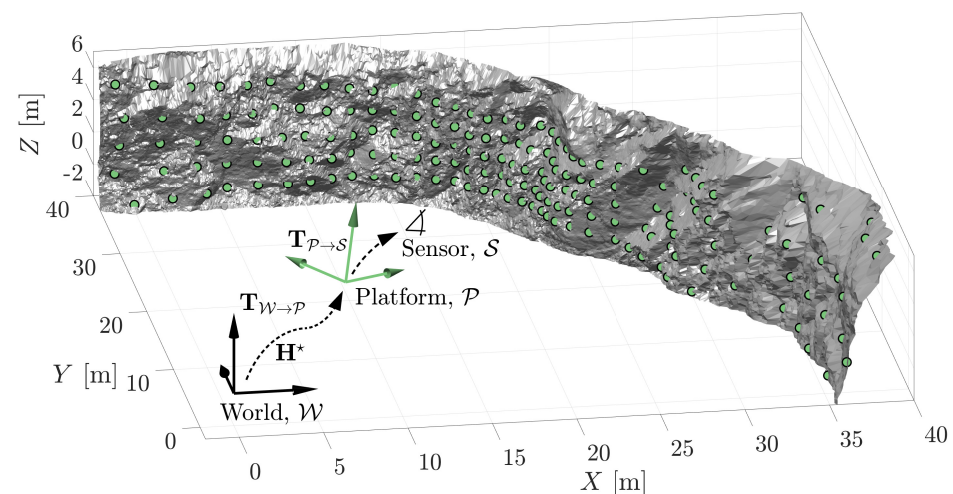
**Figure 18.** The maximum sum of evidence of a pose estimate can be used to identify which 3D geometry is most capable of reproducing the observed measurements. (a) Six chess piece geometry models. (b) The maximum sum of evidence pose ( $H^*$ ) of the *Knight* geometry as estimated against point cloud measurements of the *Queen*. (c) The maximum sum of evidence obtained from each geometry model normalised against the maximum sum of evidence obtained using the true geometry model.

The results show that, in all six cases, the correct geometry produces the highest sum of evidence. The determination of the *King* and *Queen* pieces proved to be the most difficult to differentiate from each other. The two geometries are very similar: it is possible to recreate the *Queen* measurement set from the *King* and *vice versa* (approximately 92% in each case). Not all model/measurement tests shared this symmetry. The *Pawn* geometry model could only produce 33% of the *King* point cloud MSOE, whereas the *King* geometry could obtain 79% of the *Pawn* point cloud's MSOE. This asymmetry occurs because the *pawn* is much smaller than the *king* and can only reproduce a fraction of the measurements.

#### 4.3. Localisation: Answering 'Where Am I?'-Type Questions

The ICP algorithm is frequently applied in Simultaneous Localisation and Mapping (SLAM) to estimate the incremental scan-to-scan frame transformation [22,23]. We demonstrate the MSOE algorithm's capability towards localisation of an already established map, however, this map could be replaced with that of a previous scan. In this way, we are effectively inverting the 'where am I?' problem into a 'where is it?' problem.

Given an a priori map of the immediate terrain described in the world frame,  $\mathcal{W}$ , the MSOE algorithm can be used to locate the pose of the platform frame,  $\mathcal{P}$ . Rather than locate a geometry model, the  $j$ -th hypothesis,  $H_j$ , locates the frame of the platform. The MSOE pose estimate,  $H_j$ , is the platform pose most likely to provide the range measurements,  $Z$ . In this way we are estimating the transform  $T_{\mathcal{W} \rightarrow \mathcal{P}}$  rather than  $T_{\mathcal{W} \rightarrow Model}$ . Figure 19 shows an estimated platform pose calculated from measurements of a 3D terrain model.



**Figure 19.** Example localisation problem. The MSOE algorithm is demonstrated to provide the pose of the platform frame,  $\mathcal{P}$ , with respect to the World coordinate frame,  $\mathcal{W}$ .

A measurement set,  $Z$ , of 185 range measurements was established from a true platform pose,  $H_{truth} = [4^\circ, 6^\circ, 37^\circ, 9 \text{ m}, 15 \text{ m}, 1 \text{ m}]$  and a fixed platform to sensor pose,  $T_{\mathcal{P} \rightarrow \mathcal{S}}$ . The MSOE estimate for the world-to-platform transformation,  $T_{\mathcal{W} \rightarrow \mathcal{P}}$ , is  $H^* = [4.02^\circ, 5.95^\circ, 37.00^\circ, 8.997 \text{ m}, 15.004 \text{ m}, 0.982 \text{ m}]$ . The origin of the estimated coordinate frame has a Euclidean translational and rotational error of 18.2 mm and  $0.054^\circ$ , respectively. Most of the error is located within the  $z$ -coordinate of the estimated platform pose. This is predictable as the terrain does not include a floor, and provides very little geometric information with which to estimate the height of the platform.

## 5. What Challenges Does MSOE Overcome?

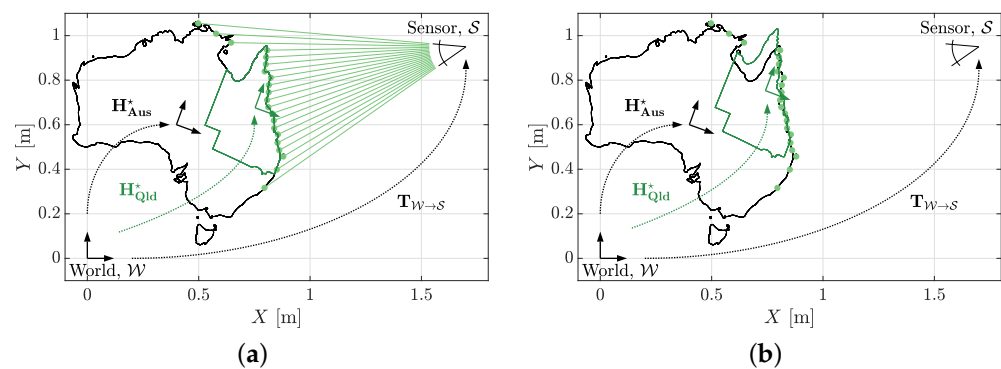
The following sections are provided to give insight into the performance of the MSOE algorithm when dealing with commonly arising challenges in pose estimation. The behaviour of the evidence-based MSOE algorithm is compared to that of the cost-based Iterative Closest Point (ICP) method. ICP is considered to be the state of the art method for object pose estimation in point cloud data.

### 5.1. MSOE Does Not Require Point Cloud Segmentation

The estimation of object pose in cluttered point clouds presents a significant challenge to algorithms that employ cost-based metrics. While segmentation routines exist, challenging environments can produce unwanted, and uncontrollable, range measurements, e.g., environments where dust is present [9].

Methods such as the Iterative Closest Point method seek to minimise the distance between the point cloud measurements and the geometry model. The problem with this approach is that it is based in the assumption that all points within the point cloud belong to the model. ICP yields excellent results when this is true, but poor results otherwise.

Consider the estimation problem illustrated in Figure 20. A point cloud has been obtained from a two dimensional shape representing the Australian mainland. Figure 20a shows the MSOE pose of both an Australian border ( $H^*_{Aus}$ ) and Queensland border ( $H^*_{Qld}$ ) estimated from the complete measurement set.

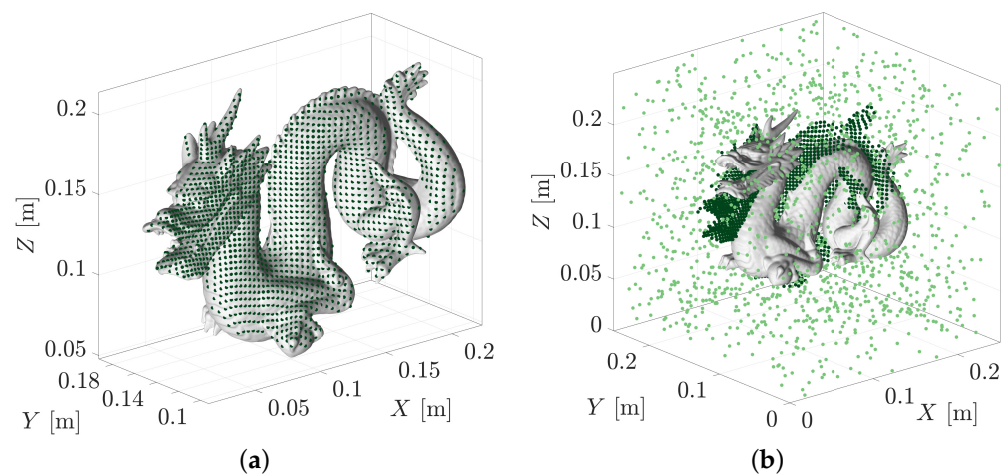


**Figure 20.** Models of Australia (black) and Queensland (green) fit to a point cloud measurement set. (a) The MSOE algorithm is able to locate both geometries, however (b) the ICP algorithm is unable to locate the Queensland model without prior segmentation of the point cloud.

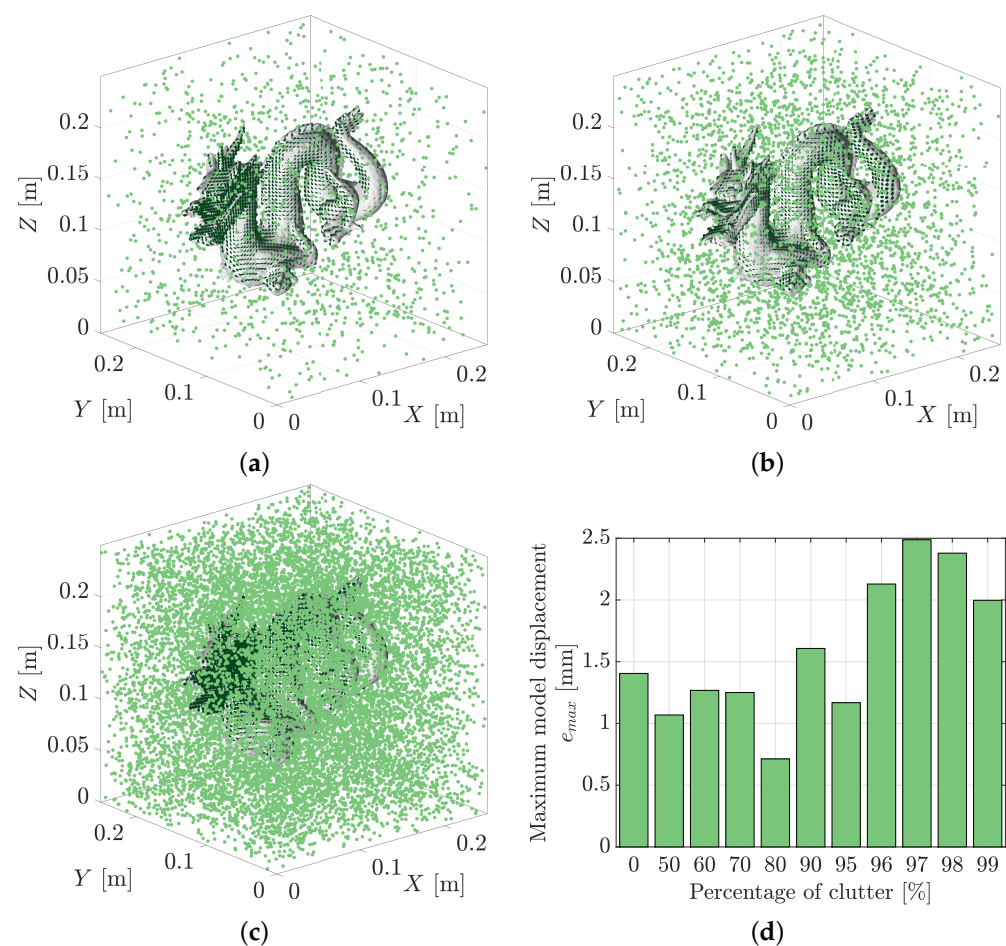
Figure 20b shows the results of this example repeated with the ICP method. The border model of Australia is correctly located, however, the estimated pose of the Queensland border is shown to be in considerable error. The failure mechanism here is similar to that observed in Section 2. The measurements that do not belong to Queensland produce an undesirable influence on the ICP pose solution. MSOE effectively ignores measurements that do not add evidence.

Figure 21a shows the ICP fit of the Stanford Dragon geometry model to a simulated LiDAR point cloud of 1554 range measurements. The pose estimate was obtained using the base ICP implementation provided by Point Cloud Library [24]. Figure 21b shows how the pose solution is corrupted when an additional 1554 measurements of uniformly distributed clutter/noise are introduced. The failure that occurs here is a result of the cost-based metric treating each point as equally important in minimising the point-to-model distances.

The evidence-based metric used by MSOE is guided by the conditional likelihoods of obtaining each measurement,  $z_i$ , under an assumed pose hypothesis,  $H_j$ . Spurious measurements may be likely under some of the tested hypotheses, however, they will not coherently support the same pose hypothesis in the way that true measurements support the correct pose. From an information perspective, they do not detract from the available information contained within the true measurement set, they simply do not provide any additional information. Figure 22 shows the MSOE pose solutions for measurement sets consisting of 50% to 90% clutter. The point clouds in this figure contain the original 1554 correct measurements and then add in an additional 1554, 3626 and 13,986 clutter measurements for 50%, 70% and 90% clutter experiments, respectively. The MSOE pose estimate is shown to remain robust to cluttered point clouds even up to the 99% clutter test which contains 153,846 clutter measurements.



**Figure 21.** The ICP point-to-plane fit is unable to determine the pose of the model in the cluttered point cloud. (a) ICP point-to-plane fit of Stanford Dragon model to segmented point cloud. (b) ICP point-to-plane fit of Stanford Dragon model to the same point cloud with an equal number of uniformly random measurements (i.e., 50% clutter).



**Figure 22.** The maximum sum of evidence solution,  $H^*$ , is plotted against the (a) 50%, (b) 70%, and (c) 90% clutter measurement sets. Further testing was conducted on measurement sets consisting of up to 99% clutter (153,846 clutter : 1554 true). (d) The maximum geometric displacement of the dragon models,  $e_{max}$ , was found to reach up to 2.5 mm in the 97% clutter test.

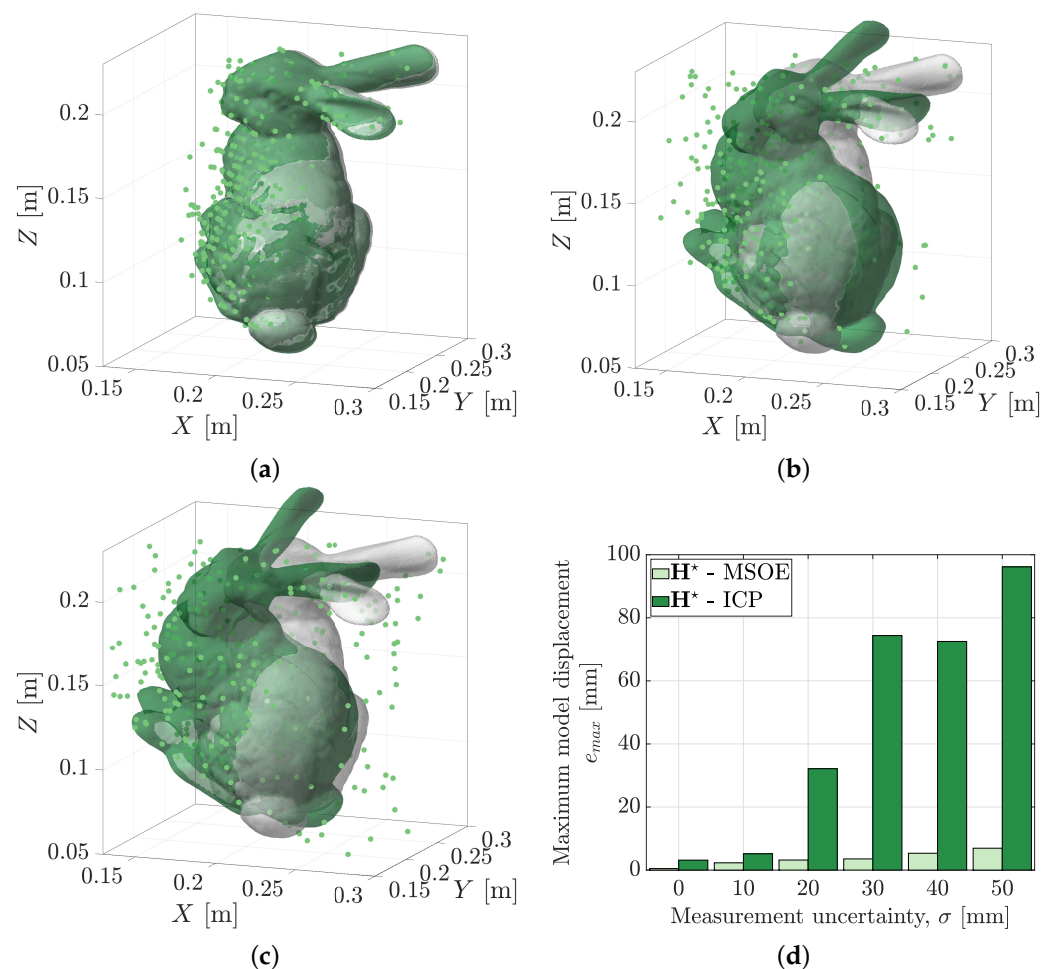
The largest observed model displacement,  $e_{max}$ , was 2.5 mm and was found to occur in the 97% noise test. Interestingly, the most accurate pose was found in the 80% noise

test which produced a solution that displaced the model by 0.7 mm (Figure 22d). This suggests that the algorithm is operating robustly when random clutter is introduced to the point cloud.

### 5.2. MSOE Is Robust to Range Measurement Uncertainty, $\sigma$

Range sensors do not provide perfect range measurements as simulated in the previous sections. Section 3.1 provided discussion on how sensor measurement uncertainty affects the range distribution (as was illustrated in Figure 6). Range sensor measurement accuracy is typically reported as having zero bias and  $\sigma^2$  variance. For example, the Velodyne VLP-16 sensor is rated to  $\sigma = 30$  mm [25], whereas survey-grade LiDAR such as the FaroFocus<sup>3D</sup> is rated to a more accurate  $\sigma = 0.3$  mm [26].

Cost-based metrics such as ICP do not lend themselves to handling measurement uncertainty, whereas MSOE is based in the range measurement space. Figure 23 shows ICP pose estimates (green) of the Stanford Bunny geometry. The solutions are estimated against range data with increasing measurement uncertainty,  $\sigma = 10$  mm to 50 mm. The pose determined from the Iterative Closest Point method is shown to rotate the geometry, aligning its major-axis with the direction of range measurements. This behaviour occurs as it is the best way to minimise the point-to-model error metric driving the solution.



**Figure 23.** ICP (green) and MSOE (grey) pose solutions obtained for simulated measurement uncertainties of: (a)  $\sigma = 10$  mm; (b)  $\sigma = 30$  mm; (c)  $\sigma = 50$  mm. (d) Maximum geometric displacement for increasing measurement uncertainty,  $\sigma$ .

Maximum sum of evidence pose estimates,  $H^*$ , are plotted in grey alongside the ICP solutions. There is no visible difference in the MSOE pose estimates as measurement



uncertainty is increased. The maximum geometric displacement,  $e_{max}$ , of the bunny geometry model is indicated for both ICP and MSOE in Figure 23d. Under the cost-based ICP approach, the maximum geometric displacement is shown to increase as high as 96.1 mm. When using the MSOE approach, the geometry model is shown to displace up to 6.9 mm (which occurred in the  $\sigma = 50$  mm test). The range measurement error used to construct this test is sampled from an unbiased Gaussian,  $\mathcal{N}(0, \sigma^2)$ . Geometric displacement of the geometry using the MSOE solutions is most likely due to any bias that occurred in the as-sampled range errors.

## 6. Conclusions and Significance

This paper applies a single evidence-based method towards the problems of pose estimation, object classification and platform localisation. The solutions to these three problem categories are shown to be accurate when assessed under the metric of maximum geometric displacement.

Poses are obtained for two dimensional and three dimensional objects and the method is shown to extend to finding solutions in other hypothesis spaces, e.g., robot joint space. The maximum sum of evidence measure is shown to provide a good classifier of objects from point cloud data. An object is better suited to a point cloud if it contains geometry that is more likely to reproduce the range measurements. Finally, MSOE is shown to provide for accurate platform localisation against a priori map geometries.

The main contribution of the MSOE approach lies in its robustness to segmentation and measurement uncertainty. This robustness is achieved by removing the need for an assumed point cloud correspondence. The algorithm's performance is tested on unsegmented data and compared to the results obtained with ICP (a cost-based method). Pose estimates are shown to be robust to unsegmented measurements of a scene as well as unsegmented random noise. MSOE is also demonstrated to find accurate pose estimates in point cloud data produced with significant range uncertainty.

The final conclusion of this paper is that the MSOE algorithm does not require any variation to handle the challenges associated with pose estimation in cluttered point clouds, e.g., segmentation. The algorithm takes unprocessed range measurements as input and provides pose estimates, object classification, or localisation as output. There are no tuning parameters required (e.g., nearest neighbour regions or voxel size). It merely requires an assumed geometry model and a representative magnitude of measurement or extrinsic calibration uncertainty in order to be configured.

We have used the MSOE algorithm in our work in Mining equipment automation. The benefits of the algorithm far outweigh the increased computational cost and we recommend its use to others who are seeking to solve 'where is it?', 'what is it?', and 'where am I?' problems.

**Author Contributions:** Conceptualisation, T.P. and P.M.; methodology, T.P.; software, T.P. and T.D.; validation, T.P.; formal analysis, T.P.; investigation, T.P. and T.D.; resources, T.P. and T.D.; data curation, T.P.; writing—original draft preparation, T.P.; writing—review and editing, T.D.; visualisation, T.P.; supervision, P.M.; project administration, T.P.; funding acquisition, T.P. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Acknowledgments:** The research was supported and conducted within the University of Queensland, St Lucia.

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



## Abbreviations

The following abbreviations are used in this manuscript:

MSOE	Maximum Sum Of Evidence
ICP	Iterative Closest Point
MISE	Mean Integrated Square Error
ISE	Integrated Square Error
KDE	Kernel Density Estimation
EM	Expectation Maximisation
LiDAR	Light Detection Furthermore, Ranging
DOF	Degree Of Freedom
PDF	Probability Density Function
CDF	Cumulative Density Function

## References

- Besl, P.J.; McKay, N.D. Method for registration of 3-D shapes. In *Sensor Fusion IV: Control Paradigms and Data Structures*; International Society for Optics and Photonics: Bellingham, WA, USA, 1992; Volume 1611, pp. 586–606.
- Simon, D.A.; Hebert, M.; Kanade, T. Real-time 3-D pose estimation using a high-speed range sensor. In Proceedings of the 1994 IEEE International Conference on Robotics and Automation, San Diego, CA, USA, 8–13 May 1994; pp. 2235–2241.
- Grundmann, T.; Eidenberger, R.; Zoellner, R.D.; Xue, Z.; Ruehl, S.; Zoellner, J.M.; Dillmann, R.; Kuehnle, J.; Verl, A. Integration of 6D object localization and obstacle detection for collision free robotic manipulation. In Proceedings of the 2008 IEEE/SICE International Symposium on System Integration, Nagoya, Japan, 4 December 2008; pp. 66–71.
- Halma, A.; ter Haar, F.; Bovenkamp, E.; Eendebak, P.; van Eekeren, A. Single spin image-ICP matching for efficient 3D object recognition. In Proceedings of the ACM Workshop on 3D Object Retrieval, Firenze, Italy, 25 October 2010; pp. 21–26.
- Paulino da Silva, J.; Lamar, M.V.; Bordim, J.L. Accuracy and efficiency performance of the ICP procedure applied to sign language recognition. *CLEI Electron. J.* **2014**, *17*, 11.
- Borrmann, D.; Elseberg, J.; Lingemann, K.; Nüchter, A.; Hertzberg, J. Globally consistent 3D mapping with scan matching. *Robot. Auton. Syst.* **2008**, *56*, 130–142. [\[CrossRef\]](#)
- Nobili, S.; Scona, R.; Caravagna, M.; Fallon, M. Overlap-based ICP tuning for robust localization of a humanoid robot. In Proceedings of the 2017 IEEE International Conference on Robotics and Automation (ICRA), Singapore, 29 May–3 June 2017; pp. 4721–4728.
- Chen, Y.; Medioni, G. Object modelling by registration of multiple range images. *Image Vis. Comput.* **1992**, *10*, 145–155. [\[CrossRef\]](#)
- Phillips, T.G.; Guenther, N.; McAree, P.R. When the dust settles: The four behaviors of lidar in the presence of fine airborne particulates. *J. Field Robot.* **2017**, *34*, 985–1009. [\[CrossRef\]](#)
- Phillips, T.G.; McAree, P.R. An evidence-based approach to object pose estimation from LiDAR measurements in challenging environments. *J. Field Robot.* **2018**, *35*, 921–936. [\[CrossRef\]](#)
- Lachhani, K.; Duan, J.; Baghsiahi, H.; Willman, E.; Selviah, D.R. Error Metric for Indoor 3D Point Cloud Registration. In *2014 Irish Machine Vision and Image Processing Conference, IMVIP 2014, Irish Pattern Recognition and Classification Society (IPRCS), a Member Body of the International Association for Pattern Recognition (IAPR)*; IPRCS Irish Pattern Recognition and Classification Society: Sligo, Ireland, 2014; Volume 2014, pp. 34–42.
- Choi, O.; Park, M.G.; Hwang, Y. Iterative K-Closest Point Algorithms for Colored Point Cloud Registration. *Sensors* **2020**, *20*, 5331. [\[CrossRef\]](#) [\[PubMed\]](#)
- D’Adamo, T.A.; Phillips, T.G.; McAree, P.R. Registration of three-dimensional scanning LiDAR sensors: An evaluation of model-based and model-free methods. *J. Field Robot.* **2018**, *35*, 1182–1200. [\[CrossRef\]](#)
- Phillips, T.G.; Green, M.E.; McAree, P.R. An adaptive structure filter for sensor registration from unstructured terrain. *J. Field Robot.* **2015**, *32*, 748–774. [\[CrossRef\]](#)
- Martinez, W.L.; Martinez, A.R. *Computational Statistics Handbook with MATLAB*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2015; Chapter 8.
- Turlach, B.A. Bandwidth selection in kernel density estimation: A review. In *CORE and Institut de Statistique*; Citeseer: University Park, PA, USA, 1993.
- Silverman, B.W. *Density Estimation for Statistics and Data Analysis*; Routledge: London, UK, 1986.
- Botev, Z.I.; Grotowski, J.F.; Kroese, D.P. Kernel density estimation via diffusion. *Ann. Stat.* **2010**, *38*, 2916–2957. [\[CrossRef\]](#)
- Li, H.; Zhang, K.; Jiang, T. *The Regularized EM Algorithm*; AAAI: Menlo Park, CA, USA, 2005; pp. 807–812.
- Sheather, S.J.; Jones, M.C. A reliable data-based bandwidth selection method for kernel density estimation. *J. R. Stat. Soc. Ser. B (Methodol.)* **1991**, *53*, 683–690. [\[CrossRef\]](#)
- Peterson, H. Why McDonald’s Chicken McNuggets Come in only 4 Shapes. Available online: <https://www.businessinsider.com/why-mcdonalds-chicken-mcnuggets-have-four-shapes-2014-2>(accessed on 17 September 2021).

22. Holz, D.; Behnke, S. Sancta simplicitas-on the efficiency and achievable results of SLAM using ICP-based incremental registration. In Proceedings of the 2010 IEEE International Conference on Robotics and Automation, Anchorage, Alaska, 3–8 May 2010; pp. 1380–1387.
23. Zhang, J.; Singh, S. LOAM: Lidar Odometry and Mapping in Real-time. In Proceedings of Robotics: Science and Systems, Berkeley, CA, USA, 12–16 July 2014 ; Volume 2.
24. Rusu, R.B.; Cousins, S. 3D is here: Point cloud library (pcl). In Proceedings of the 2011 IEEE International Conference on Robotics and Automation (ICRA), Shanghai, China, 9–13 May 2011; pp. 1–4.
25. Velodyne. *Lidar Vlp-16 User Manual Furthermore, Programming Guide*; Velodyne: San Jose, CA, USA, 2016.
26. FARO Technologies. *FARO Laser Scanner Focus3D Manual*; FARO Technologies: Lake Mary, FL, USA, 2013.