Towards Intensity-Augmented SLAM with LiDAR and ToF Sensors

Robert A. Hewitt[1] and Joshua A. Marshall[2]

Abstract— Although passive sensors are widely used for many mobile robotics applications that perform mapping and localization functions, there are many environments (e.g., mining and planetary) where active sensors are more practical. However, at present, most 3D SLAM algorithms that do use LiDAR and/or time-of-flight (ToF) sensors exploit only range and bearing information associated with these measurements, but not intensity information. This paper presents a new approach that attempts to explicitly incorporate an intensity model as part of a sparse bundle adjustment (SBA) estimation problem. An observability analysis shows that a solution exists, and simulation results verify its potential utility.

I. INTRODUCTION

All LiDAR-based SLAM algorithms use the range data that LiDAR data provides, but very few make use of the intensity data that accompanies each range measurement. So far, the few who have used intensity measurements predominantly use them as a way to improve upon the popular Iterative Closest Point (ICP) algorithm [1] or form intensity images and apply appearance-based computer vision techniques to describe and detect 2D features and match them across images [2]. However, if intensity—like range measurements—could be modelled and predicted based on the robot and map state variables, as well as known sensor noise characteristics, this would open up the possibility of using intensity as an additional measurement to constrain the estimates of robot pose and the position of landmarks in its environment. It could also lead to the addition of reflectivity and surface normals as estimated parameters in the robot's map.

The primary motivation for the work presented in this paper is to develop accurate and computationally efficient LiDAR-based SLAM algorithms for potential deployment on ESA's planetary exploration rovers, while also envisioning the application of these results to Earth applications (e.g., in mining) where robotic vehicles work in environments having unpredictable lighting conditions or no external lighting.

Firstly, we ask whether it is possible to model intensity measurements as a function of robot and map parameters. The answer to this question lies in a large body of previous work. In the computer vision literature the problem of *shape-from-shading* bears many similarities to the problem at hand.

By assuming a known light source direction and modelling brightness as a function of surface shape, it has been shown that shape can be derived from intensity gradients in camera images [3]. As an extension to this, varying known light source directions across multiple images of the same scene, has been used in a technique called *photometric stereo* to estimate surface shape and reflectivity [4].

Höfle and Pfeifer [5] described a method for producing arial LiDAR intensity images that are proportional to the reflectance of the measured surface by modelling the power return of the laser as a diffuse reflection and estimating the surface normal using nearest neighbor range measurements. modelling non-diffuse reflections in a scene viewed by passive sensors has been demonstrated in the computer graphics domain [6] with parametric models that contain both diffuse and specular parts in a given reflection. Thus, despite the complexity of modelling light reflections, information can be recovered about the scene from camera intensity images. A common assumption is a known light source direction, which has limited the use of these techniques in robotics. However, in the case of active sensors this is usually true.

Can additional state parameters in the map provide a benefit to the applications for which SLAM is envisaged? Assuming that surface normal and reflectivity of laser data could be accurately estimated, there are examples in the literature of using this information with classifiers for both engineering and geological applications. Wurm et al. [7] were able to discern vegetation from streets based on LiDAR intensity, range and incidence angle information. Tatoglu and Pochiraju used intensity measurements to classify the reflective properties of scanned surfaces [8] by testing how well the intensity matched various reflective models such as Lambertian or Phong models.

Burton et al. [9] showed in geological outcrop studies that intensity has a log-linear correlation to shale and sand percentage . This relationship was used to estimate important properties of un-weathered outcrops. Another potential use of reflectivity data is performing science operations on a planetary rover. The Mars Science Laboratory (MSL) has been able to detect hydrated minerals at a distance by analyzing the intensity of infrared and visual spectrum images taken by its mast camera [10]. These same techniques could be applied to LiDAR reflectivity maps.

Does adding these measurements improve the observability, accuracy, and/or speed of SLAM? This question is of core interest to the European Space Agency (ESA), where improved localization with active sensors is desired in anticipation of varying lighting conditions that future rover missions will encounter in shadowed regions of the

[1]R. A. Hewitt is with the Department of Electrical and Computer Engineering, Queen's University, Kingston, Ontario, Canada and is currently conducting research at the Autonomous Robotics Laboratory at the European Space Research and Technology Centre `rob.hewitt@esa.int`

[2]J. A. Marshall is with the Department of Electrical and Computer Engineering and the Robert M. Buchan Department of Mining, Queen's University, Kingston, Ontario, Canada `joshua.marshall@queensu.ca`

Moon or Mars missions that conduct traverses over the course of an entire day. In this paper, we show that by introducing intensity measurements the problem requires less landmarks than a typical bundle adjustment problem and in a simulated environment improves upon the accuracy and speed of computing a solution.

II. INTENSITY MEASUREMENTS

LiDAR and ToF cameras emit light (from a laser or LEDs) and measure the returning light that is reflected from surrounding objects to determine the range between the target and the sensor. In scanning LiDAR the azimuth and elevation of the lighting source is also estimated by a combination of calibration, encoders, and/or an IMU. In ToF cameras, azimuth and elevation are replaced by pixel values. The sensor also measures the power or amplitude of the returning light, and compares this with the power or amplitude at which it was emitted. The ratio between the two is often called *intensity* by active sensor manufacturers and is usually reported alongside range measurements.

The physical principles that determine the power of the returning signal are the same as microwave radar [5]. The radar range equation comprises the three main factors: (a) the sensor, (b) the target, and (c) atmospheric parameters. An assumption here is the target surface is roughly Lambertian, which means that light is scattered evenly in all directions when reflected. For the emitted signal with power P_e,

$$P_r = \frac{P_e D_r^2 \rho}{4R^2} \eta_{\text{sys}} \eta_{\text{atm}} \cos \alpha, \qquad (1)$$

where P_r is the returning power of the signal, D_r is the receiver aperture diameter, ρ is the target reflectivity, R is the range to the target, η_{sys} is the system transmission factor, η_{atm} is the atmospheric transmission factor and α is the incident angle. The intensity measurement is linearly proportional to P_r by some constant parameters of the LiDAR sensor, η_{LiDAR}, that can be grouped with other constant parameters η_{sys}, η_{atm}, D_r, and P_e to form η_{all};

$$I = \eta_{\text{LiDAR}} P_r \qquad (2)$$

$$= \eta_{\text{LiDAR}} \frac{P_e D_r^2 \rho}{4R^2} \eta_{\text{sys}} \eta_{\text{atm}} \cos \alpha \qquad (3)$$

$$= \frac{\rho}{R^2} \eta_{\text{all}} \cos \alpha. \qquad (4)$$

Looking at (1), we can see that P_r is a function of the range, which itself is a function of the target position and the sensor pose, the incidence angle which is determined by the sensor orientation and the surface normal, and on ρ (the target's reflectivity), which depends on the material of the target.

III. PROBLEM FORMULATION

A. Notation and Conventions

The approach described in this paper requires the use of quaternions [11], surface normal vectors, matrices, and scalars. This section describes their notation, required operations, as well as other relevant properties of these quantities.
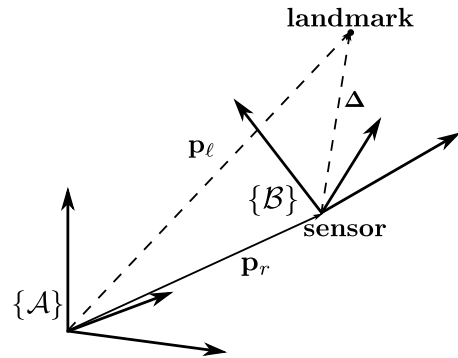

Fig. 1: Coordinate frames for 3D localization of sensor and landmarks in the global frame

1) Coordinate Frames: Two coordinate frames are defined to express the position and orientation of the vehicle and landmarks. Coordinate frame $\{\mathcal{A}\}$ is the global frame and coordinate frame $\{\mathcal{B}\}$ is attached to the sensor. It is assumed here that $\{\mathcal{B}\}$ represents both the LiDAR and vehicle coordinate frames. A unit quaternion $\mathbf{q} = (\mathbf{q}_v, q_s) \in \mathbb{S}^3$, is used to represent the rotation between these frames where subscripts v and s denote the vector and scalar parts of the quaternion, respectively. Positions are described by a 3-vector $\mathbf{p} = (p_x, p_y, p_z) \in \mathbb{R}^3$. The transformation between frames is depicted in Fig. 1.

2) Quaternion Notation: Unit quaternions exist on a unit 3-sphere in $\mathbb{R}^4$, which is denoted $\mathbb{S}^3$. Hence, they must satisfy the constraint $\mathbf{q}^T \mathbf{q} = 1$. Quaternions are described by a scalar value q_s and a vector $\mathbf{q}_v$; i.e.,

$$\mathbf{q} = \begin{bmatrix} \mathbf{q}_v \\ q_s \end{bmatrix} = \begin{bmatrix} q_x & q_y & q_z & q_s \end{bmatrix}^\top. \qquad (5)$$

Quaternion multiplication is performed according to

$$\mathbf{q}_c = \mathbf{q}_a \otimes \mathbf{q}_b = \begin{bmatrix} q_{a_s} \mathbf{q}_{b_v} + q_{b_s} \mathbf{q}_{a_v} + \mathbf{q}_{a_v} \times \mathbf{q}_{b_v} \\ q_{a_s} q_{b_s} - \mathbf{q}_{a_v} \cdot \mathbf{q}_{b_v} \end{bmatrix}, \qquad (6)$$

which is non-commutative due to the cross product term. The rotation from $\{\mathcal{A}\}$ to $\{\mathcal{B}\}$ is denoted ${}^{\mathcal{B}}_{\mathcal{A}}\mathbf{q}$ and a vector ${}^{\mathcal{A}}\mathbf{p}$ in $\{\mathcal{A}\}$ can be rotated to $\{\mathcal{B}\}$ by

$$\mathbf{q}_{{}^{\mathcal{B}}\mathbf{p}} = {}^{\mathcal{B}}_{\mathcal{A}}\mathbf{q} \otimes \mathbf{q}_{{}^{\mathcal{A}}\mathbf{p}} \otimes {}^{\mathcal{B}}_{\mathcal{A}}\mathbf{q}^{-1}, \qquad (7)$$

where $\mathbf{q_p} = [\mathbf{p}^\top, 0]^\top$ and, because the inverse of a unit quaternion is equal to its conjugate, $\mathbf{q}^{-1} = [-\mathbf{q}_v^\top, q_s]^\top$.

3) Normal Vectors and Landmarks: Given a 3D surface and a point on that surface, a surface normal vector is simply the vector that is perpendicular to the tangent plane to the surface at that point. The light that is reflected from a surface is dependent on the angle between the light source direction and this surface normal.

Normal vectors live on a unit 2-sphere embedded in $\mathbb{R}^3$, denoted $\mathbb{S}^2$. Surface normals are not unique; for a surface that exists in $\mathbb{R}^3$, at any given point on that surface there is an inward and outward pointing surface normal. However, in the case of active sensor measurements it can be assumed that the surface normal points towards the sensor. Surface normals, described by the 3-vector $\mathbf{n} = (n_x, n_y, n_z) \in \mathbb{S}^2$ with $||\mathbf{n}|| =$
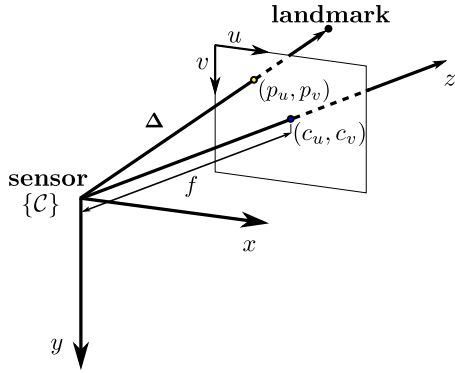
Fig. 2: Camera reference frame $\{\mathcal{C}\}$ and a pinhole camera image plane

1, are used in addition to reflectivity, ρ, and position $\mathbf{p}$, to parameterize, in this paper, what we refer to as *landmarks* $\boldsymbol{\ell} = \left(p_{\ell_x}, p_{\ell_y}, p_{\ell_z}, n_x, n_y, n_z, \rho\right) \in \mathbb{R}^3 \times \mathbb{S}^2 \times [0, 1]$.

B. Intensity-Augmented SLAM

The technique presented here is a form of SLAM that estimates m sensor poses, $\mathbf{s} = ((\mathbf{p}_{r,1}, \mathbf{q}_1), ..., (\mathbf{p}_{r,m}, \mathbf{q}_m))$ that are each associated with a set of sensor observations of the same set of static landmarks $\boldsymbol{\ell} = (\boldsymbol{\ell}_1, ..., \boldsymbol{\ell}_n)$, which include intensity-based information, all in global frame $\{\mathcal{A}\}$. For the purposes of this paper, no other knowledge or observations (such as the motion between measurements) is used and landmark recognition—also called *data association*—for each set of measurements is assumed to be known.

The goal of this SLAM algorithm is, thus, to estimate the state vector that describes the sensor pose and each landmark state in the global frame $\{\mathcal{A}\}$, given all observations. The total state vector for the problem that includes all sensor poses and landmark positions is $\mathbf{x} = (\mathbf{s}_1, ..., \mathbf{s}_m, \boldsymbol{\ell}_1, ..., \boldsymbol{\ell}_n)$. Let a subset of the state vector, $\mathbf{x}_{i,j}$, represent the i-th sensor pose and j^{th} feature parameters $\mathbf{x}_{i,j} = (\mathbf{s}_i, \boldsymbol{\ell}_j)$.

C. Measurement Models

Scanning LiDAR measurements are usually obtained by using a photodiode to directly measure the time of flight of a laser pulse and its amplitude. Depending on how the laser and photodiode are mounted on the sensor their angle of elevation is usually constant and their angle of azimuth is measured with an encoder. With the measured time of flight, the distance the light travelled can be computed and the sensor typically reports the return as range r, azimuth β and elevation α. ToF cameras work on a different principle, in which a series of LEDs bathe the scene in light modulated at a known frequency, and the onboard CCD pixel array absorbs the reflected light and measures the phase difference. Due to the known wavelength of the modulated signal, the range can be computed from this phase difference. The ToF camera data are typically reported as range r and pixel locations described by p_u and p_v as shown in Fig. 2. As long as the image has been corrected for any spherical or tangential distortions, the pixel locations in a 2D image are equivalent to the elevation and bearing angles in scanning LiDAR measurements. For brevity's sake, only the model for a scanning LiDAR measurement is included here.

An observation of the j-th landmark at the i-th LiDAR scanner pose can be modelled as

$$\mathbf{h}_s(\mathbf{x}_{i,j}) = \begin{bmatrix} r_{i,j} \\ \beta_{i,j} \\ \alpha_{i,j} \end{bmatrix} = \begin{bmatrix} \sqrt{{}^{\mathcal{B}}\Delta_{x_{i,j}}^2 + {}^{\mathcal{B}}\Delta_{y_{i,j}}^2 + {}^{\mathcal{B}}\Delta_{z_{i,j}}^2} \\ \operatorname{atan}({}^{\mathcal{B}}\Delta_{y_{i,j}}, {}^{\mathcal{B}}\Delta_{x_{i,j}}) \\ \operatorname{atan}\left({}^{\mathcal{B}}\Delta_{z_{i,j}}, \sqrt{{}^{\mathcal{B}}\Delta_{x_{i,j}}^2 + {}^{\mathcal{B}}\Delta_{y_{i,j}}^2}\right) \end{bmatrix} \tag{8}$$

where

$$^{\mathcal{A}}\boldsymbol{\Delta}_{i,j} = \begin{bmatrix} {}^{\mathcal{A}}\Delta_{x_{i,j}} \\ {}^{\mathcal{A}}\Delta_{y_{i,j}} \\ {}^{\mathcal{A}}\Delta_{z_{i,j}} \end{bmatrix} = \begin{bmatrix} p_{\ell_x,j} - p_{r_x,i} \\ p_{\ell_y,j} - p_{r_y,i} \\ p_{\ell_z,j} - p_{r_z,i} \end{bmatrix}, \tag{9}$$

and

$$\begin{bmatrix} {}^{\mathcal{B}}\boldsymbol{\Delta}_{i,j} \\ 0 \end{bmatrix} = {}^{\mathcal{B}}_{\mathcal{A}}\mathbf{q}_j \otimes \begin{bmatrix} {}^{\mathcal{A}}\boldsymbol{\Delta}_{i,j} \\ 0 \end{bmatrix} \otimes {}^{\mathcal{B}}_{\mathcal{A}}\mathbf{q}_j^{-1}, \tag{10}$$

where ${}^{\mathcal{B}}_{\mathcal{A}}\mathbf{q}_j = \mathbf{q}_j^{-1}$.

This paper seeks to incorporate additional measurements in the SLAM algorithm; intensity I and a measurement, $\check{\mathbf{n}}$, of the surface normal. Information about the orientation of these surfaces is lost in the sampling process, however the information still exists in the relationship between a sampled point and points in its neighbourhood [12]. The literature provides two predominant methods for estimating the surface normal of a point and its uncertainty; one based on optimization methods, and one based on averaging methods [12]. While optimization methods generally provide better results, averaging methods are less computationally complex. In both cases, the surface normal is measured in the sensor frame and therefore the measurement model is a function of the sensor orientation and the surface normal estimates.

In this paper, our measurement model assumes Lambertian reflectance at the surface. More general reflectance functions exist that can be used to model other types of reflection (e.g., specular) but these are not addressed in the analysis done here. We model the j-th additional measurements taken at the i-th sensor pose as

$$\begin{bmatrix} I_{i,j} \\ \check{\mathbf{n}}_{i,j} \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{\eta_{all}\rho_j}{r_{i,j}^2} \left(\mathbf{n}_j \cdot \frac{-{}^{\mathcal{A}}\boldsymbol{\Delta}_{i,j}}{\|{}^{\mathcal{A}}\boldsymbol{\Delta}_{i,j}\|} \right) \\ {}^{\mathcal{B}}_{\mathcal{A}}\mathbf{q}_i \otimes \begin{bmatrix} \mathbf{n}_j \\ 0 \end{bmatrix} \otimes {}^{\mathcal{B}}_{\mathcal{A}}\mathbf{q}_i^{-1} \end{bmatrix} \in [0, \infty) \times \mathbb{S}^2. \tag{11}$$

Thus, the measurement model that augments these additional measurements to the scanning LiDAR model is $\mathbf{h}_{sa}(\mathbf{x}_{i,j}) = (r_{i,j}, \beta_{i,j}, \alpha_{i,j}, I_{i,j}, \check{\mathbf{n}}_{i,j})$.

IV. OBSERVABILITY ANALYSIS

This section provides an observability analysis of a three dimensional SLAM problem with a scanning LiDAR landmark sensor that measures range, bearing, elevation, intensity, and includes a surface normal estimate, taken at different sensor poses. First, the system's observability when incorporating only range, bearing, and elevation measurements is analyzed, with intensity and surface normal measurements added to investigate how this affects observability.

A sufficient condition of local weak observability [13] is that the observability rank condition is satisfied. Local observability is a stronger condition than global observability [14]. The system is locally weakly observable if it is "possible to instantaneously distinguish (the states) from their neighbors (in the local state space) for all possible states". Our SLAM problem satisfies this condition if any of the observability matrices are of rank $6m + 3n$ for traditional SLAM, and rank $6m + 6n$ with SLAM that includes reflectivity and surface normal estimates (Section III-B). The traditional SLAM problem analysis includes only the position of landmarks and is estimated with the measurements in (8). A similar analysis, using the augmented state, is conducted and the observability of both is compared. The approach used here has been extended from [14], which is about 2D SLAM and from [15] which extends these concepts to 3D SLAM.

For a nonlinear function, such as $\mathbf{h}(\mathbf{x})$, the observability matrix is formed such that the elements in this matrix are any combination of the repeated Lie derivatives of the components of the measurement model $\mathbf{h}(\mathbf{x})$, and the process model $\mathbf{f}(\mathbf{x}, \mathbf{u})$ that describes how the system states change with time given an input $\mathbf{u}$. An observability matrix is composed by using any combination of successive Lie derivatives of any of the components of the measurement model. The Lie derivatives are defined recursively as

$$L_f^0 dh(\mathbf{x}) = \frac{\partial \mathbf{h}(\mathbf{x})}{\partial \mathbf{x}}, \tag{12}$$

$$L_f^d dh(\mathbf{x}) = L_f^{d-1} dh(\mathbf{x}) \frac{\partial \mathbf{f}(\mathbf{x}, \mathbf{u})}{\partial \mathbf{x}} \tag{13}$$
$$+ \left[\frac{\partial}{\partial \mathbf{x}} \left(L_f^{d-1} dh(\mathbf{x}) \right)^T \mathbf{f}(\mathbf{x}, \mathbf{u}) \right]^T.$$

Because the system described here treats each state as static, $\mathbf{f} = \mathbf{0}_{(6m+3n) \times 1}$, only the first Lie derivative (i.e., the Jacobian of the measurement model) is relevant. However for a system that incorporates measurements of motion (e.g., an IMU), further Lie derivatives would be computed. For traditional SLAM, the observability matrix is simply any combination of $6m + 3n$ rows that are taken from any of the successive Lie derivatives. If any of these possible observability matrices can be shown to be full rank, then the system is locally weakly observable.

For example, a system with only one landmark ($n = 1$) and one sensor pose ($m = 1$) has a state vector with 9 parameters and 3 observations. A potential observability matrix for this system is

$$\mathbf{O} = \begin{bmatrix} L_F^0 dh_1(\mathbf{x}) \\ L_F^0 dh_2(\mathbf{x}) \\ L_F^0 dh_3(\mathbf{x}) \\ \vdots \\ L_F^2 dh_1(\mathbf{x}) \\ L_F^2 dh_2(\mathbf{x}) \\ L_F^2 dh_3(\mathbf{x}) \\ L_F^3 dh_1(\mathbf{x}) \end{bmatrix}, \tag{14}$$

where h_k refers to the k-th row of the measurement vector. However, because only the first Lie derivative is non-trivial, we can see right away that this system is unobservable and has at most a rank equal to 3. To determine the minimum number of estimated sensor poses and landmarks needed to construct an observability matrix we can set up the inequality

$$6m + 3n \leq 3mn, \tag{15}$$

which shows that this always requires at least two sensor poses. For two sensor poses, we require at least four landmarks to satisfy inequality (15) and for any number of poses greater than two we require at least three landmarks. With two sensor poses and four landmarks, only one observability matrix exists that includes all possible Lie derivatives

$$\mathbf{O} = \left(\mathbf{O}_{1,1}, \mathbf{O}_{1,2}, ..., \mathbf{O}_{2,4} \right), \tag{16}$$

where

$$\mathbf{O}_{i,j} = \begin{bmatrix} L_F^0 dh_1(\mathbf{x}_{i,j}) \\ L_F^0 dh_2(\mathbf{x}_{i,j}) \\ L_F^0 dh_3(\mathbf{x}_{i,j}) \end{bmatrix}, i = 1, \ldots 2 \text{ and } j = 1, \ldots, 4. \tag{17}$$

We determined the rank of $\mathbf{O}$ by using MATLAB's Symbolic Toolbox, which implements Gaussian elimination. In the case of two robot poses ($m = 2$) and four landmarks ($n = 4$), $\mathbf{O}$ must have a rank of $6m + 3n = 24$. However, carrying out the analysis it can be shown that rank($\mathbf{O}$) = 18 which is a rank deficiency of six. This is a well known problem in the SLAM community [14], and it can be shown that no matter how many landmarks or sensor poses there are, $\mathbf{O}$ is always rank deficient. This is due to the inability to observe the global frame, $\{\mathcal{A}\}$. To alleviate this, the initial sensor pose can be fixed at the origin of $\{\mathcal{A}\}$. This reduces the minimum rank requirement to $6(m-1)+3n$ and carrying out the same operations reveals that the problem is indeed locally weakly observable.

Using the augmented measurement model that includes intensity and nearest neighbor surface normal measurements, there are now additional landmark states to estimate that have three degrees of freedom ($\mathbf{n} \in \mathbb{S}^2$ and $\rho \in [0,1]$). Here we devise three types of "augmented" SLAM that all still make use of the standard LiDAR measurements but also use augmented measurements: (1) one that uses only the first augmented measurement I and estimates each landmark's associated $\mathbf{n}$ and ρ; (2) one that only uses the second augmented measurement, $\check{\mathbf{n}}$, and just estimates $\mathbf{n}$; and (3) one that uses both augmented measurements and estimates $\mathbf{n}$ and ρ. The Lie derivatives are computed as before to construct the observability matrix. We can determine the minimum number of estimated sensor poses and landmarks needed to construct an observability matrix by the inequality

$$6m + 6n \leq 6nm. \tag{18}$$

Thus, at minimum the problem requires two sensor poses and two landmarks. Immediately we see a difference from traditional SLAM, with only half the landmarks needed to construct the observability matrix. The rank deficiency for

Measurements	Dim. of ℓ	n for 2 views	n for 3 views
$\mathbf{h} = (r, \beta, \alpha)$	$3n$	4(2)	3(2)
$\mathbf{h} = (r, \beta, \alpha, I)$	$6n$	6(3)	3(2)
$\mathbf{h} = (r, \beta, \alpha, \check{\mathbf{n}})$	$5n$	3(2)	2(2)
$\mathbf{h} = (r, \beta, \alpha, I, \check{\mathbf{n}})$	$6n$	2(1)	2(1)

TABLE II
Rank deficiency with minimum number of landmarks,
brackets indicate value when first frame is fixed

Measurements	2 views	3 views
$\mathbf{h} = (r, \beta, \alpha)$	6(0)	6(0)
$\mathbf{h} = (r, \beta, \alpha, I)$	12(6)	6(0)
$\mathbf{h} = (r, \beta, \alpha, \check{\mathbf{n}})$	3(0)	3(0)
$\mathbf{h} = (r, \beta, \alpha, I, \check{\mathbf{n}})$	3(0)	3(0)

Fig. 3: Simulation with five 6D robot poses and 20 land-marks; orientation and surface normal direction are indicated

TABLE III
Standard deviation (STD) of simulated sensor
measurements

Measurement Type	STD (σ)
Range, (σ_r)	0.01 m
Bearing, (σ_β)	$0.5°$
Elevation, (σ_α)	$0.5°$
Intensity, (σ_I)	164
Unit surface normal component, (σ_n)	0.03

this problem can be shown to be three, a reduction from six in traditional SLAM with four landmarks. If the initial robot pose is fixed to the origin of $\{\mathcal{A}\}$, the problem becomes fully observable. This has some encouraging properties over traditional SLAM. We are now estimating the shape and infrared reflectivity of the landmarks in our map, useful properties for scientific analysis and landmark recognition. Additionally, the problem is observable with fewer landmarks, which means it may perform better when the number of visible landmarks is reduced, which is typical for low resolution sensors like scanning LiDAR and ToF cameras.

The conditions needed to construct an observability matrix with two or three separate view points for the sensor are shown in Table I for all types of SLAM discussed previously. The rank deficiency of the problem is shown for the same set of circumstances in Table II.

Interestingly, the minimum number of landmarks needed for the different types of measurement vectors is the last case, which includes all measurements. By fixing the first frame, an observability matrix with just one landmark can be constructed.

As shown in Table II, the SLAM problem becomes full rank when fixing the first frame in all cases, except when only augmenting the measurement vector with intensity from two viewpoints. In summary, by estimating properties of the surface normals and reflectivity, SLAM is observable under less strict requirements and we gain additional useful information about the surface normal and infrared reflectivity of each landmark compared to traditional SLAM.

V. SIMULATION

Sparse bundle adjustment (SBA) as described by Lourakis and Argyros [16] was implemented using MATLAB to verify that the proposed method of SLAM using augmented intensity and nearest neighbor surface normal measurements was valid. The simulation was tested with known data association and for each sensor pose all landmarks were visible and
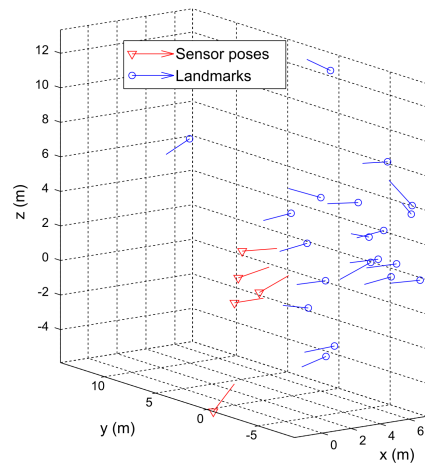
measured. The first sensor pose estimate was fixed to the origin of $\{\mathcal{A}\}$. The true pose of the first sensor was also set to the origin of $\{\mathcal{A}\}$ for ease of plotting.

The rest of the true sensor poses and landmark values were randomly perturbed such that the camera poses were near to the origin and pointing along the global x-axis, while the landmarks were pointing in the opposite direction (again, perturbed) and were randomly distributed along a plane approximately 7 m away. For the tests presented here, five sensor poses and 20 landmarks were estimated. An example of a random set of true values is shown in Fig. 3.

To compare using augmented measurements to traditional SLAM the algorithm was run with and without intensity and surface normal measurements. The measurement standard deviation (STD) is shown in Table III, all modelled as $\sim \mathcal{N}(0, \sigma^2)$. The intensity noise properties are not available from the manufacturers, so a value of 1% of the maximum possible value was chosen as a reasonable amount based on the STD observed over hundreds of measurements of a static scene by active sensors like the MESA Imaging SR4000 ToF camera and the Velodyne HDL-32E LiDAR. The surface normal noise was added to each simulated measurement and then the surface normal vector was normalised to maintain it as a unit vector. The intrinsic sensor parameter, η_{all}, was treated as a known constant, however this must be determined experimentally. Alternatively, η_{all} could be estimated on-line as an additional parameter to account for changes in atmosphere and temperature with time.

TABLE IV
Simulation Results for 100 runs

State	Initial RMSE	Traditional SLAM RMSE	Augmented SLAM RMSE
$\mathbf{p}_r$	$(1.0\,,\,1.0\,,\,1.0\,)$ m	$(0.093, 0.088, 0.11)$ m	$(0.034, 0.024, 0.021)$ m
$\mathbf{q}$	$(5.0, 5.0, 5.0)^\circ$	$(0.13, 0.11, 0.11)^\circ$	$(0.16, 0.11, 0.13)^\circ$
$\mathbf{p}_\ell$	$(2.0, 2.0, 2.0)$ m	$(0.10, 0.091, 0.11)$ m	$(0.034, 0.041, 0.038)$ m
$\mathbf{n}$	$(0.20, 0.20, 0.20)$ m	n/a	$(0.021, 0.020, 0.019)$ m
ρ	0.20	n/a	0.037

The estimates of sensor pose and landmark parameters were initialized with a root mean squared error (RMSE) that is shown in Table IV. After running the SBA algorithm the results show that it is possible to estimate all parameters in the problem using the augmented measurements. Note that $\mathbf{q}$ values have been converted to an Euler angle representation. Over 100 simulations, the RMSE of the robot pose estimates and landmark position estimates were smaller when using the augmented measurements. The additional landmark parameters were also estimated within an acceptable amount of error. The full results of these tests are shown in Table IV. In addition to these results, the mean run time of the augmented SLAM algorithm on an i7 desktop computer was found to be approximately 8% faster than the traditional SLAM algorithm due to less iterations until convergence.

VI. DISCUSSION AND CONCLUSION

The work presented in this paper shows that, under certain assumptions, the addition of intensity measurements from LiDAR and ToF cameras has promise for improving the performance of SLAM applications that use these sensors. By including intensity-based information, the SLAM problem can be solved with fewer landmarks. Alternatively, with the same number of landmarks the solution is more accurate.

However, there remain several practical considerations to address. This work currently relies on repeatedly-measured landmarks that reflect infrared light diffusely. In real-world environments, objects often include specular reflections that are directional. One could use outlier detection, such as RANSAC [17], to detect non-Lambertian reflecting landmarks that do not agree with the model and not use them. In fact, because only one landmark is needed to solve for the transformation between two robot poses, outlier detection can be computed much faster than with only range measurements. Along these lines specular components of reflection can be removed before the returning light is measured by using polarization filters [18]. Another solution is to use more advanced models that include specular reflections. These options are being investigated by the authors to bring this technique from simulation and into the field.

ACKNOWLEDGMENTS

REFERENCES

[1] G. Godin, M. Rioux, and R. Baribeau, "Three-dimensional registration using range and intensity information," in *Proceedings of SPIE: Videometrics III*, vol. 2350, (Boston, Massachusetts, U.S.A.), pp. 279–290, International Society for Optics and Photonics, 1994.

[2] C. McManus, P. Furgale, B. Stenning, and T. D. Barfoot, "Lighting-invariant visual teach and repeat using appearance-based lidar," *Journal of Field Robotics*, vol. 30, no. 2, pp. 254–287, 2013.

[3] R. Zhang, P.-S. Tsai, J. E. Cryer, and M. Shah, "Shape-from-shading: a survey," *Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 8, pp. 690–706, 1999.

[4] R. Basri, D. Jacobs, and I. Kemelmacher, "Photometric stereo with general, unknown lighting," *International Journal of Computer Vision*, vol. 72, no. 3, pp. 239–257, 2007.

[5] B. Höfle and N. Pfeifer, "Correction of laser scanning intensity data: Data and model-driven approaches," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 62, pp. 415–433, Dec. 2007.

[6] B. T. Phong, "Illumination for computer generated pictures," *Communications of the ACM*, vol. 18, pp. 311–317, June 1975.

[7] K. M. Wurm, H. Kretzschmar, R. Kümmerle, C. Stachniss, and W. Burgard, "Identifying vegetation from laser data in structured outdoor environments," *Robotics and Autonomous Systems*, Nov. 2012.

[8] A. Tatoglu and K. Pochiraju, "Point cloud segmentation with LIDAR reflection intensity behavior," in *Proceedings of the International Conference on Robotics and Automation*, (Saint Paul, Minnesota, U.S.A.), pp. 786–790, IEEE, May 2012.

[9] D. Burton, D. B. Dunlap, L. J. Wood, and P. P. Flaig, "Lidar intensity as a remote sensor of rock properties," *Journal of Sedimentary Research*, vol. 81, pp. 339–347, May 2011.

[10] M. Rice, E. Cloutis, J. Bell, D. Bish, B. Horgan, S. Mertzman, M. Craig, R. Renaut, B. Gautason, and B. Mountain, "Reflectance spectra diversity of silica-rich materials: Sensitivity to environment and implications for detections on Mars," *Icarus*, vol. 223, pp. 499–533, Mar. 2013.

[11] S. L. Altmann, *Rotations, quaternions, and double groups*. Mineola, New York, U.S.A.: Dover Publications, Inc., 2005.

[12] K. Klasing, D. Althoff, D. Wollherr, and M. Buss, "Comparison of surface normal estimation methods for range sensing applications," in *Proceedings of the International Conference on Robotics and Automation*, (Kobe, Japan), pp. 3206–3211, IEEE, May 2009.

[13] R. Hermann and A. Krener, "Nonlinear controllability and observability," *IEEE Transactions on Automatic Control*, vol. 22, pp. 728–740, Oct. 1977.

[14] K. Lee, W. Wijesoma, and I. Javier, "On the observability and observability analysis of SLAM," in *Proceedings of the International Conference on Intelligent Robots and Systems*, (Beijing, China), pp. 3569–3574, IEEE, Oct. 2006.

[15] F. Aghili, "3D simultaneous localization and mapping using IMU and its observability analysis," *Robotica*, vol. 29, pp. 805–814, Oct. 2011.

[16] M. I. A. Lourakis and A. A. Argyros, "SBA: A software package for generic sparse bundle adjustment," *ACM Transactions on Mathematical Software*, vol. 36, pp. 1–30, Mar. 2009.

[17] M. Fischler and R. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, 1981.

[18] L. B. Wolff, "Using polarization to separate reflection components," in *Proceedings of the Conference on Computer Vision and Pattern Recognition*, (San Diego, California, U.S.A.), pp. 363–369, IEEE, 1989.