2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 25-29, 2020, Las Vegas, NV, USA (Virtual)

A Theory of Fermat Paths for 3D Imaging Sonar Reconstruction

Eric Westman, Ioannis Gkioulekas, and Michael Kaess

***Abstract*— In this work, we present a novel method for reconstructing particular 3D surface points using an imaging sonar sensor. We derive the two-dimensional Fermat flow equation, which may be applied to the planes defined by each discrete azimuth angle in the sonar image. We show that the Fermat flow equation applies to boundary points and surface points which correspond to specular reflections within the 2D plane defined by their azimuth angle measurement. The Fermat flow equation can be used to resolve the 2D location of these surface points within the plane, and therefore also their full 3D location. This is achieved by translating the sensor to estimate the spatial gradient of the range measurement. This method does not rely on the precise image intensity values or the reflectivity of the imaged surface to solve for the surface point locations. We demonstrate the effectiveness of our proposed method by reconstructing 3D object points on both simulated and real-world datasets.**

I. Introduction

Imaging sonars, or forward-looking sonars (FLS), have been widely used as a core sensing modality for autonomous underwater vehicles (AUVs). They prove particularly useful in turbid waters, in which the range of optical cameras is severely limited [14]. Mapping underwater scenes or structures with imaging sonars has been a topic of increasing interest, as rich 3D maps are crucial to many autonomous tasks. The focus of this work is advancing the state-of-the-art of 3D reconstruction with imaging sonars.

In many ways, the imaging sonar sensor is analogous to the monocular optical camera. Both sensors provide 2D measurements of a 3D environment. Just as the optical camera measures azimuth and elevation angles, but loses range in perspective projection, so the imaging sonar measures azimuth and range, but not elevation. This is visualized in Fig. 1a. Recovering the elevation angle of imaged surfaces is therefore the fundamental challenge of 3D imaging sonar reconstruction.

A variety of previous works attempt to reconstruct the 3D location of sparse feature points by using an extended Kalman filter (EKF) [13, 25] or nonlinear optimization [9, 10, 11, 15, 21, 22]. The evolutionary algorithm CMA-ES has also been used to address the nonlinearity of the problem and to avoid getting stuck in local minima [6]. These methods focus on minimizing the reprojection error of feature points that are corresponded across multiple images. One of the main drawbacks of this general approach is the difficulty of reliably and automatically detecting and corresponding feature points from different sonar viewpoints. Many of these works rely on manually extracted features or simulated features and correspondences. More recently, A-KAZE features [1] have been used for real-time feature detection and correspondence. However, correspondence is usually only feasible between nearby sonar viewpoints, limiting the variety of constraints that can be obtained on the feature points [12, 16, 21].

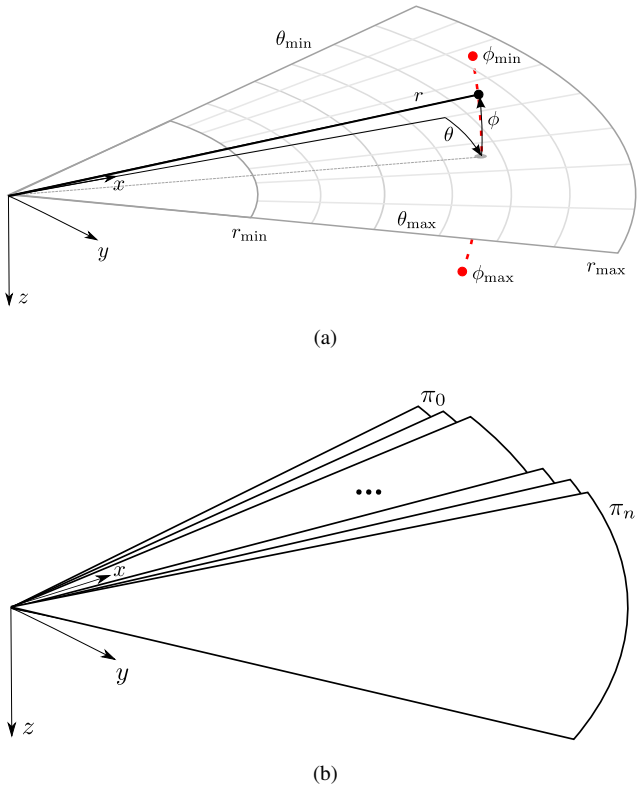

Fig. 1: (a) The imaging sonar sensor model. The coordinate frame of the sensor is defined with the x-axis pointing forward and z-axis pointing downward. An imaged 3D point (black dot) is projected into the zero-elevation plane, where it is imaged by the pixel corresponding to its azimuth angle θ and range r. (b) The ith image column corresponds to a 2D plane π_i which contains the sensor's z-axis.

Other efforts have been made to generate dense object models using imaging sonar. Generative models have been used to create a dense pointcloud surface estimate for each sonar image [2, 4], which may be fused from multiple viewpoints in a truncated signed distance field (TSDF) to create a single cohesive model [20]. However, these methods require one or more 3D estimates of object edges and knowledge of the reflectivity properties of the imaged surfaces, both of which are difficult to obtain in real-world environments. Vol-

This work was partially supported by the Office of Naval Research under grant N00014-16-1-2103 and by the DARPA REVEAL program under contract HR0011-16-C-0025.

The authors are with the Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15213, USA. {westman, igkioule, kaess}@cmu.edu

umetric methods such as occupancy grid mapping [18, 19] or albedo-based models [8, 23] create a discretized volume of the scene and use filtering or optimization to determine which voxels contributed to the observed image measurements. The principle of space carving [3, 5, 7, 8] utilizes only the shortest range high-intensity measurement at each azimuthal angle to "carve out" free space. Capturing images from a wide variety of viewpoints generates a surface or volume model that, with no noise, completely encompasses the imaged object. However, space carving requires a wealth of different viewpoints in order to generate a useful object model, which may not be feasible to obtain in the field. Additionally, while a convex object may be reconstructed with arbitrary accuracy given sufficient sonar viewpoints, there are a variety of concave geometries that space carving is incapable of reconstructing.

In this work, we present a novel method for reconstructing particular points on imaged surfaces, which we call *Fermat paths*. We derive this from the fundamental geometric principles of the imaging sonar sensor. In contrast to previous methods, our proposed framework cannot be categorized as inherently sparse or dense – the density of reconstructed points depends on the geometry of the imaged surfaces relative to the sensor. Our proposed method may be viewed as an improvement on and generalization of space carving, and is largely based on the seminal theory of Fermat paths for non-line-of-sight reconstruction [24].

The specific contributions of this work are:

- relating the problems of imaging sonar reconstruction and non-line-of-sight reconstruction,
- a derivation of the 2D Fermat flow equation, which describes a novel method for estimating the 3D location of surface points,
- a framework for estimating the spatial gradients of Fermat pathlengths to solve the 2D Fermat flow equation,
- an evaluation of the proposed method on simulated and experimental datasets.

The remainder of this paper is organized as follows: Section II describes the fundamentals of the imaging sonar sensor and analyzes relevant properties of Fermat paths for imaging sonar reconstruction. The 2D Fermat flow equation is derived in Section III as well as a method for solving it to reconstruct 3D points. Section IV presents results of our proposed method on simulated and real-world datasets and Section V offers concluding remarks.

II. Fermat paths for imaging sonar

A. Imaging sonar sensor model

The imaging sonar is an active acoustic sensor that emits pulsed sound waves and measures the intensity of sound reflected back towards it. The known speed of sound in water is used to measure the range r of returning sound waves and a 1-D array of transducers is utilized to disambiguate the azimuth angle θ. This results in a two-dimensional image where the rows correspond to discrete range bins and columns to discrete azimuth angle bins. Fig. 1a shows the image pixels projected onto the zero-elevation plane, and an imaged 3D point projected into its corresponding pixel. The sensor has a limited field of view in azimuth, range, and elevation. A single sensor origin is considered the source of emitted sound and the point of detection of reflected sound, with the x-axis pointing forward towards the imaged volume and z-axis pointed downward.

Since the sensor disambiguates the azimuth angle θ, we consider the image formation model for an arbitrary azimuth angle θ_i. Although each azimuthal beam images a slice of volume with a small non-zero width, we can approximate the image formation model in two dimensions. We consider the 2D plane π_i corresponding to azimuthal beam θ_i, which intersects the sensor's z-axis and is rotated off of the xz plane by θ_i, as shown in Fig. 1b. The rest of the discussion on Fermat paths takes place within the 2D plane. Then, using $\mathcal{X}$ to denote the 2D cross-section of the 3D scene defined by π_i, $\mathbf{c}$ to denote the imaging sonar origin, and $\mathbf{x}$ to denote a surface point along $\mathcal{X}$, the image formation model for column i may be expressed as

$$I_i\left(r;\mathbf{c}\right) = \int_{\mathcal{X}} f\left(\mathbf{x};\mathbf{c}\right)\delta\left(r - r\left(\mathbf{x};\mathbf{c}\right)\right)d\mathbf{x} \tag{1}$$

where $f\left(\mathbf{x};\mathbf{c}\right)$ captures visibility, reflectivity, shading, and the spatial propagation of sound. $r\left(\mathbf{x};\mathbf{c}\right)$ denotes the range of a surface point $\mathbf{x}$ from the sensor origin.

For the sake of simplicity, we consider only single bounce returns, disregarding that sound waves may be reflected off of multiple surfaces before returning to the sensor, falsely contributing intensity to pixels at longer ranges. Additionally, we define a coordinate system in the plane π_i as follows: the z-axis is aligned with the sensor's z-axis, and the forward-facing w-axis lies in the sensor's xy-plane, rotated from the x-axis by θ_i.

B. Fermat paths in imaging sonar sensing

Our theory follows the work of Xin et al. [24], which derives the 3D Fermat flow equation for non-line-of-sight (NLOS) reconstruction. Under this model, recovering a point's elevation angle is equivalent to 2D range-only mapping. This is akin to the problem of NLOS reconstruction, which may be viewed as 3D range-only mapping, as the azimuth angle is not disambiguated.

Consider a scene $\mathcal{X}$ that is formed as the union of smooth surfaces. This theory focuses on a subset of surface points, defined as follows, which possess unique properties that make them amenable to 3D reconstruction.

Definition 1. *Let $\mathbf{x} \in \mathcal{X}$ be a point on the scene surface, $\mathbf{c}$ the sonar sensor origin, and $r\left(\mathbf{x};\mathbf{c}\right) = \|\mathbf{x} - \mathbf{c}\|$ the range (or pathlength) of the surface point $\mathbf{x}$ from the sensor. Then, the Fermat set $\mathcal{F}\left(\mathbf{c}\right)$ is the set of all points $\mathbf{x}$ for which the range function $r\left(\mathbf{x};\mathbf{c}\right)$ is a local extremum or a saddle point.*

We refer to points in $\mathcal{F}\left(\mathbf{c}\right)$ as Fermat paths because they correspond to paths that satisfy Fermat's principle.

Proposition 2. *The Fermat set consists of two disjoint sets, the boundary set $\mathcal{B}\left(\mathbf{c}\right)$ and the specular set $\mathcal{S}\left(\mathbf{c}\right)$, such that $\mathcal{F}\left(\mathbf{c}\right) \triangleq \mathcal{B}\left(\mathbf{c}\right) \cup \mathcal{S}\left(\mathbf{c}\right)$. $\mathcal{B}\left(\mathbf{c}\right)$ contains all points on $\mathcal{X}$ for*

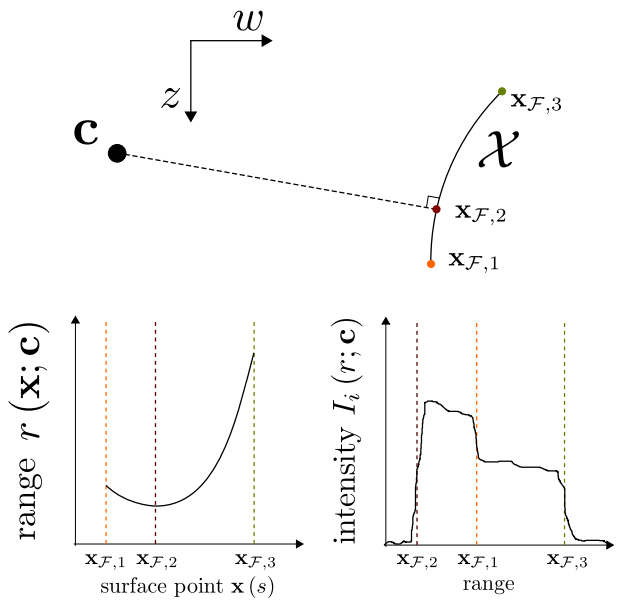

Fig. 2: A surface $\mathcal{X}$ observed by an imaging sonar located at $\mathbf{c}$ in the 2D plane corresponding to a particular azimuth angle. The first return is generated by $\mathbf{x}_{\mathcal{F},2}$, which is the only point in the specular set in this example. The two boundary points are $\mathbf{x}_{\mathcal{F},1}$ and $\mathbf{x}_{\mathcal{F},3}$. Below the drawing, the range measurement is plotted as a function of the surface point and the image intensity is shown as a function of the range.

which a normal vector is not defined. $\mathcal{S}(\mathbf{c})$ *consists of points that create a specular reflection.*

Proof: Let $s \in [0,1]$ be a parameterization of the object surface $\mathcal{X}$. By fundamental principles, the range function $r(\mathbf{x}(s);\mathbf{c})$, has extrema at $s=0$, $s=1$. Other extrema or saddle points only occur where $\frac{\partial r(\mathbf{x}(s);\mathbf{c})}{\partial s}=0$. Boundary points correspond to $s=0$ and $s=1$ and are thus clearly extrema. For a specular point $\mathbf{x}(s) \in \mathcal{S}(\mathbf{c})$, consider the derivative of the range:

$$\frac{\partial r(\mathbf{x}(s);\mathbf{c})}{\partial s} = \left\langle \frac{\mathbf{x}(s)-\mathbf{c}}{\|\mathbf{x}(s)-\mathbf{c}\|}, \mathbf{x}_s(s) \right\rangle \tag{2}$$

$$= \frac{1}{r(\mathbf{x}(s);\mathbf{c})} \langle \mathbf{x}(s)-\mathbf{c}, \mathbf{x}_s(s) \rangle \tag{3}$$

where $\mathbf{x}_s(s) = \frac{\partial \mathbf{x}(s)}{\partial s}$, and $\|\cdot\|$ denotes the ℓ_2-norm. The vector $\mathbf{x}_s(s)$ is by definition parallel to the tangent of the curve $\mathcal{X}$ at $\mathbf{x}$. The surface normal at $\mathbf{x}(s)$ must be parallel or anti-parallel to $\mathbf{x}(s)-\mathbf{c}$ in order to create a specular reflection. Therefore, $\mathbf{x}(s)-\mathbf{c}$ is orthogonal to the tangent vector, so $\langle \mathbf{x}(s)-\mathbf{c}, \mathbf{x}_s(s) \rangle = 0$ and

$$\frac{\partial r(\mathbf{x}(s);\mathbf{c})}{\partial s} = 0. \tag{4}$$

Therefore, $\mathbf{x}(s)$ is either a local extremum or saddle point in the range function. ■

A simple example is shown in Fig. 2. A convex surface is imaged, resulting in three Fermat paths. The two boundary points correspond to extrema in the range from the sensor. The first-detected surface is the specular point $\mathbf{x}_{\mathcal{F},2}$, which is a local minimum in the range function.

Note that points in the specular set are not necessarily points of specular reflection in 3D. The orthogonality condition need only hold in the 2D cross section of the 3D surface defined by the azimuthal plane of interest.

Next we describe how Fermat paths may be detected in a sonar image.

Proposition 3. *Assume that the reflection of sound off the surface* $\mathcal{X}$ *is non-zero in the specular direction. Then, for all* $\mathbf{x} \in \mathcal{F}(\mathbf{c})$*, the image intensity* $I_i(r;\mathbf{c})$ *will have a discontinuity at pathlength* $r(\mathbf{x};\mathbf{c})$.

For brevity, we omit the full proof and refer the reader to [24] for the proof in the 3D scenario, which also applies directly to our 2D scenario. However, Section III-C will provide some insight into how the proof works in the boundary case. In particular, the Fermat flow constraint can be obtained by differentiating Eqn. 15 for $j=1$ point, and assuming that b is independent of sensor location.

For specular points, consider the example in Fig. 2. For a convex surface, if the shortest-range measurement is not a boundary point, it will be a specular point, as is $\mathbf{x}_{\mathcal{F},2}$. For a noiseless sensor, this will result in a jump from zero-intensity to some non-zero intensity at the corresponding range. This generalizes to specular points on concave portions of the surface, but the jump is from higher intensity to lower intensity.

With this observation, we have now identified the Fermat paths and noted that they can be detected in individual sonar image columns as discontinuities in the intensity values.

III. Fermat Flow Equation

A. Fermat flow derivation

Here we derive the Fermat flow equation, which is stated as follows and can be used to solve for the 3D locations of Fermat paths.

Proposition 4. *Consider a range measurement* $r_{\mathcal{F}}(\mathbf{c})$ *corresponding to a Fermat path. Assume that there is a single unique point* $\mathbf{x}_{\mathcal{F}} \in \mathcal{F}(\mathbf{c})$ *with* $r(\mathbf{x}_{\mathcal{F}};\mathbf{c}) = r_{\mathcal{F}}(\mathbf{c})$ *. Then,*

$$\mathbf{x}_{\mathcal{F}} = \mathbf{c} - r_{\mathcal{F}} \nabla_{\mathbf{c}} r_{\mathcal{F}}(\mathbf{c}). \tag{5}$$

Proof. We prove this for a point $\mathbf{x}_{\mathcal{F}} \in \mathcal{S}(\mathbf{c})$ in the specular set, and omit the proof for the boundary set for brevity. We will use $\mathbf{v} = [v^w, v^z]^T$ to denote the 2D coordinates of a point $\mathbf{v}$ in the plane. Let $s \in [0,1]$ be a parameterization of the surface $\mathcal{X}$ in the neighborhood around $\mathbf{x}_{\mathcal{F}}$, so that $\mathbf{x}_{\mathcal{F}} = \mathbf{x}(s(\mathbf{c}))$. Then, considering the derivative of $r_{\mathcal{F}}(\mathbf{c})$ with respect to c^w, we have

$$\frac{\partial r_{\mathcal{F}}(\mathbf{c})}{\partial c^w} = \frac{\partial \|\mathbf{x}_{\mathcal{F}}-\mathbf{c}\|}{\partial c^w} \tag{6}$$

$$= \left\langle \frac{\mathbf{x}_{\mathcal{F}}-\mathbf{c}}{\|\mathbf{x}_{\mathcal{F}}-\mathbf{c}\|}, \frac{\partial(\mathbf{x}_{\mathcal{F}}-\mathbf{c})}{\partial c^w} \right\rangle \tag{7}$$

$$= \left\langle \frac{\mathbf{x}_{\mathcal{F}}-\mathbf{c}}{\|\mathbf{x}_{\mathcal{F}}-\mathbf{c}\|}, \frac{\partial(\mathbf{x}(s(\mathbf{c}))-\mathbf{c})}{\partial c^w} \right\rangle \tag{8}$$

$$= \left\langle \frac{\mathbf{x}_{\mathcal{F}}-\mathbf{c}}{\|\mathbf{x}_{\mathcal{F}}-\mathbf{c}\|}, \mathbf{x}_s(s(\mathbf{c})) \frac{\partial s(\mathbf{c})}{\partial c^w} - [1,0]^T \right\rangle \tag{9}$$

where $\mathbf{x}_s(s(\mathbf{c}))$ is tangent to the surface $\mathcal{X}$ at $\mathbf{x}_{\mathcal{F}}$ by definition. Since $\mathbf{x}_{\mathcal{F}} \in \mathcal{S}(\mathbf{c})$ and from the definition of the

specular set $\mathcal{S}(\mathbf{c})$, the vector $\mathbf{x}_\mathcal{F} - \mathbf{c}$ is parallel to the surface normal at $\mathbf{x}_\mathcal{F}$. Therefore, $\mathbf{x}_\mathcal{F} - \mathbf{c}$ is orthogonal to $\mathbf{x}_s(s(\mathbf{c}))$, and (9) becomes

$$\frac{\partial r_\mathcal{F}(\mathbf{c})}{\partial c^w} = -\frac{(\mathbf{x}_\mathcal{F} - \mathbf{c})^w}{\|\mathbf{x}_\mathcal{F} - \mathbf{c}\|}. \tag{10}$$

The same derivation may be used to show that (10) holds for the z-coordinate as well, and therefore that

$$\nabla_\mathbf{c} r_\mathcal{F}(\mathbf{c}) = -\frac{\mathbf{x}_\mathcal{F} - \mathbf{c}}{\|\mathbf{x}_\mathcal{F} - \mathbf{c}\|}. \tag{11}$$

This may be rearranged in the form of (5). ■

The significance of this result is that Fermat paths may be reconstructed in the 2D plane, and therefore in 3D, with "single-shot" estimation. Only the gradient of the Fermat pathlength is required, which can be estimated with short baseline motion. This yields what may be thought of as a "differential stereo" algorithm for reconstructing points in $\mathcal{F}(\mathbf{c})$.

B. Fermat flow estimation

An interesting result of (11) is that the gradient of the Fermat pathlength is a unit vector, with just one degree of freedom. This gradient with respect to the acoustic center of the sensor may be computed as

$$\nabla_\mathbf{c} r_\mathcal{F}(\mathbf{c}) = \left(\sqrt{1 - \left(\frac{\partial r_\mathcal{F}}{\partial z}\right)^2}, \frac{\partial r_\mathcal{F}}{\partial z} \right)\Bigg|_\mathbf{c} \tag{12}$$

where, $\frac{\partial r_\mathcal{F}}{\partial z}$ is the gradient along the sensor's z-axis, estimated by translating the sensor in that direction. The direction of the gradient in the plane's w-axis is inferred, using the unit norm property of $\nabla_\mathbf{c} r_\mathcal{F}(\mathbf{c})$. Alternatively, $\frac{\partial r_\mathcal{F}}{\partial w}$ could be estimated by translating in plane's the w-axis. However, since the 3D direction of the w-axis for each azimuthal plane in the sonar image is different, each translational direction would yield a direct gradient estimation for only one image column. Additionally, as the sensor translates, the surface point imaged by the discontinuity in each image column may change due to translation outside the plane. Therefore, it is best to translate the sensor along its z-axis so that the azimuthal plane for each column does not change, and directly estimate $\frac{\partial r_\mathcal{F}}{\partial z}$ for all image columns simultaneously.

To generate a gradient estimate that is robust to noise in the Fermat pathlength measurement and the sensor pose, we fit a quadratic polynomial to the Fermat pathlength measurements $r_\mathcal{F}$ as a function of z:

$$r_\mathcal{F}(z) = a_2 z^2 + a_1 z + a_0 \tag{13}$$

and compute the gradient as

$$\frac{\partial r_\mathcal{F}}{\partial z} = 2a_2 z + a_1. \tag{14}$$

This smoothing procedure is applied to a window of local values around each point of interest. Note that this is a heuristic used to provide robust gradient estimates in the presence of noise, and a variety of other filtering techniques could be used instead. Fitting a quadratic to the Fermat pathlength can also be viewed as placing a prior on the smoothness or curvature of the surface.

C. Boundary points

A significant distinction between points in $\mathcal{B}(\mathbf{c})$ and those in $\mathcal{S}(\mathbf{c})$ is that in the 2D plane, $\mathcal{B}(\mathbf{c})$ consists of only up to two points for each continuous surface for all sensor locations $\mathbf{c}$. These same boundary points in $\mathcal{B}(\mathbf{c})$ are observed repeatedly from different viewpoints and correspondence may be established across these viewpoints. This is in contrast to points in $\mathcal{S}(\mathbf{c})$, of which there are infinitely many and which cannot be corresponded between different viewpoints. If the same point $\mathbf{b} \in \mathcal{B}(\mathbf{c})$ is observed from multiple sensor locations, a nonlinear least squares optimization may be used to estimate its location in the plane:

$$\min_\mathbf{b} \sum_j \left\| r_\mathbf{b}(\mathbf{c}_j) - \sqrt{\left(b^w - c_j^w\right)^2 + \left(b^z - c_j^z\right)^2} \right\|^2 \tag{15}$$

where $r_\mathbf{b}(\mathbf{c}_j)$ is the range measurement of the point from sensor location $\mathbf{c}_j$.

This estimation method may be thought of as a 2D equivalent of the nonlinear optimizations previously used to solve for the 3D locations of sparse features under general sensor motion [9, 15, 22]. The main drawback of these previous methods is reliably detecting and corresponding feature points between different viewpoints. However, restricting the sensor motion to translation in the z-axis provides an implicit solution to this problem: boundary points may be detected as discontinuities in the image and tracked within the same image column as the sensor moves. However, discontinuities corresponding to boundary points can generally not be distinguished from discontinuities corresponding to specular points, without the aid of prior information.

D. Field of view

In our derivation of the Fermat flow equation, we disregard the sensor's field of view, assuming that either the sensor has an unlimited elevation field of view or that the entire surface $\mathcal{X}$ lies within the sensor's frustum. In reality, imaging sonars have a limited elevation field of view. For example, the SoundMetrics DIDSON[1] can be configured for up to 28° elevation aperture and the Oculus M-series[2] sonars for up to 20°. If a surface lies partially within the sensor's field of view, our analysis remains intact and visible specular and boundary points may still be reconstructed using the Fermat flow equation. The one complicating effect is that the intersection of the surface with the end of the sensor's field of view in the elevation direction may generate additional discontinuities in the image that correspond neither to specular points nor boundary points. If the Fermat flow equation is applied to all discontinuities in a sequence of sonar images, the presence of such a geometry could potentially introduce false estimated surface points, since the Fermat flow equation does not hold for these points.

[1] www.soundmetrics.com/Products/DIDSON-Sonars/DIDSON-300m

[2] www.blueprintsubsea.com/oculus/index.php

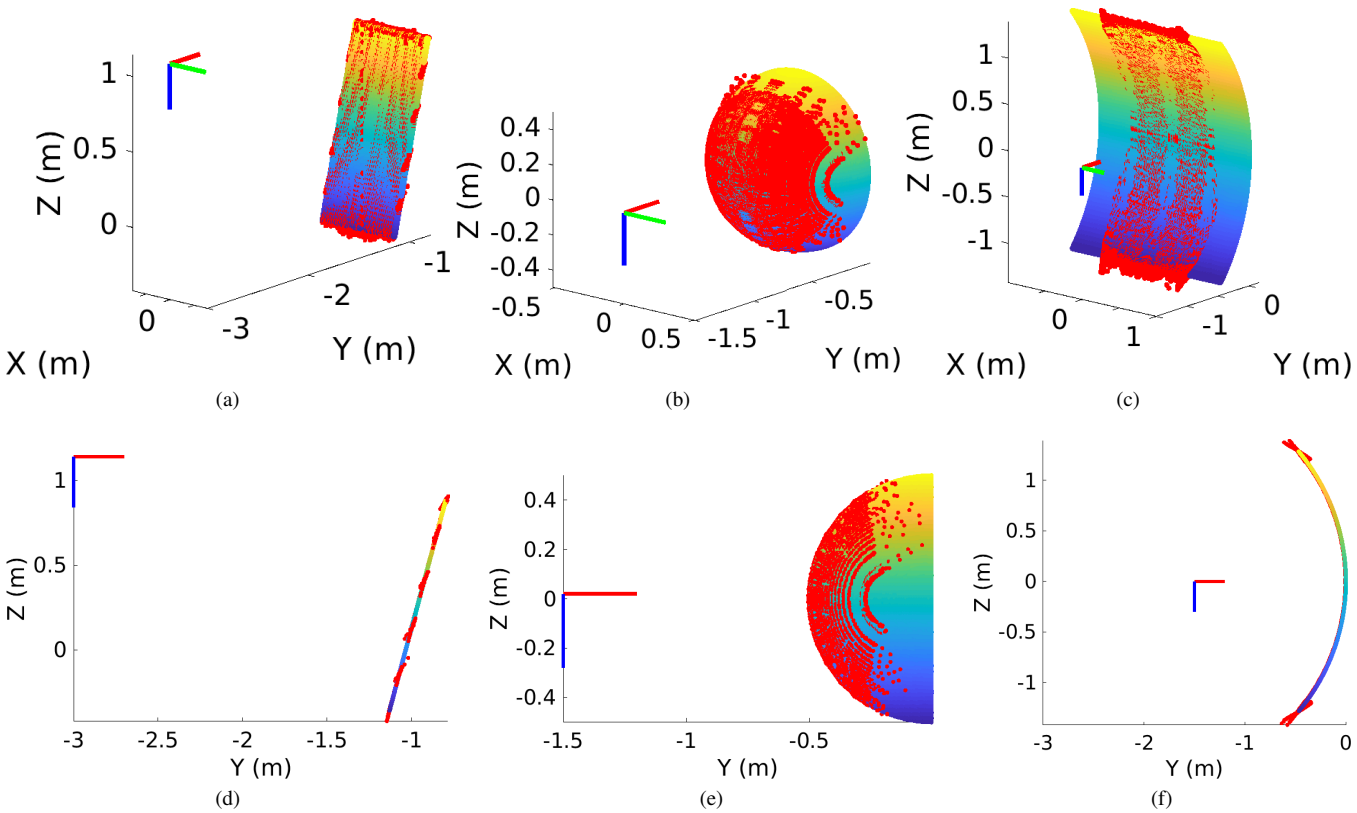

Fig. 3: Reconstructed points of surfaces from simulated datasets using our solution to the Fermat flow equation. (a) - (c) depict isometric views of the ground truth surface points (yellow to blue colored) and the reconstructed surface points (red) for sections of a plane, sphere, and concave cylinder, respectively. (d) - (f) depict side views of the same reconstructions.

E. Relation to space carving

The application of the Fermat flow equation to reconstruct specular points may be viewed as a generalization of the theory of space carving for imaging sonar reconstruction [3, 5]. Space carving considers the first high-intensity return for each image column, which is generated by the closest surface in the 2D plane defined by the azimuthal beam. All volume along the entire elevation field of view between the sensor and this range measurement is "carved out" as free space, while all volume behind this measurement remains potentially occupied. While the closest high-intensity measurement is generally caused by one surface point, the surface point is not explicitly determined or solved for. The estimated surfaces are simply the exterior of the potentially occupied regions that remain after repeating the carving process from a variety of viewpoints.

The point generating the first return is, by definition, one of the many Fermat paths of a surface. When the surface is convex (e.g., a cylinder, sphere, or plane), all of its Fermat paths will also be points generating first returns. In this case, it can be shown that, assuming an infinite density of sensors, our method and space carving will generate the same surface. However, when a surface is *sufficiently* concave, it will contain Fermat paths that do not generate first returns. In this case, space carving will generate a hull containing the surface within its interior, but will fail to reconstruct its concave parts. By contrast, our method will successfully reconstruct the entire surface. This relationship is discussed, in the context of NLOS imaging, in [17, 24].

IV. Evaluation

We present results from simulation and real-world datasets to demonstrate and evaluate the proposed framework for imaging sonar reconstruction. For both simulated and real-world sonar images, we detect discontinuities using Canny edge detection and use several heuristics to reject false positives due to noise.

A. Simulation

We generate simulated datasets of a variety of simple surfaces. We simulate an imaging sonar with an artificially wide elevation aperture of $180°$ for two purposes. First, a wider elevation aperture allows for great coverage of surfaces during a single sweep along the z-axis. Second, this also demonstrates that wider elevation apertures do not affect the accuracy of our proposed algorithm, as long as the discontinuities are detectable in the image. This contrasts starkly with volumetric albedo [8, 23] or occupancy grid [18, 19] methods which perform significantly worse with wider elevation apertures. We use a simple projective image formation model that does not model occlusions or shading as the precise pixel intensity is not used to solve the Fermat flow equation. Pixel intensities are only used for detecting discontinuities that correspond to points in the specular or boundary sets.

Figure 3 visualizes the reconstructed specular and boundary points compared to the ground truth surface models. The specular points are reconstructed nearly perfectly while there is a small amount of noise in the boundary point estimation, due to the smoothness prior enforced in estimating the Fermat pathlength gradient.

B. Real-world experiments

We also conducted real-world experiments using a Bluefin HAUV[3] robotic test platform in a test tank environment. Data was collected using a SoundMetrics DIDSON imaging sonar with a spreader lens to increase the elevation aperture to 28°. The vehicle's odometry, which makes use of a high-end IMU and Doppler velocity log (DVL), is used for sensor pose estimates at each sonar frame.

We image a planar plywood target, as shown in Fig. 4, with the sonar oriented nearly perpendicular to the target. A single sweep along the sonar's z-axis allows for reconstructing a dense set of points on along the surface and boundary of the target. The boundaries of the square cut-outs are not easily detected due to the perpendicular viewing angle. Some surface points are reconstructed in these gaps due to the violation of the assumption in Proposition 4 that a detected discontinuity is due to a single surface point – some discontinuities are the result of simultaneous returns from the plane above and below the cut-out. Nevertheless, the side view of the reconstruction demonstrates that the elevation angle of the surface points is estimated quite accurately and the planar structure is recovered despite making no assumptions regarding planarity of the scene. As expected, many noisier boundary points are reconstructed due to repeated observation. The mean absolute error of the reconstructed points is 1.3 cm, which is mostly due to the noisy estimation of boundary points. The range resolution of the sonar in this experiment is approximately 0.5 cm per pixel. Fig. 5 shows detected boundary points reconstructed using Eqn. 15, with the plane imaged at a steeper angle to aid the detection of the interior boundaries.

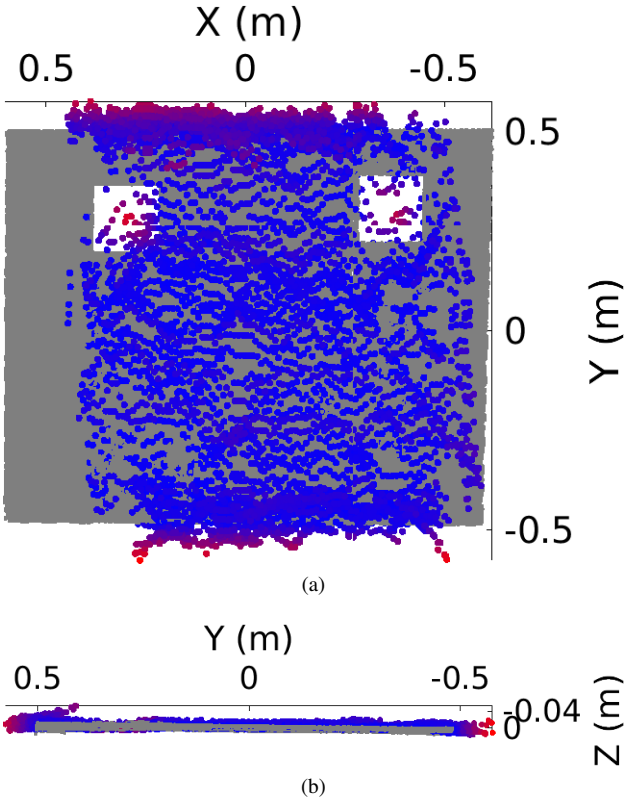

Fig. 4: (a) Top-down view and (b) side view of a planar target imaged in a test tank. The ground-truth point cloud (gray) is shown alongside the reconstructed surface points for comparison, which are colored blue for lower error and red for higher error. The reconstruction is incomplete due to the limited azimuthal field of view of the sonar (28.8°).

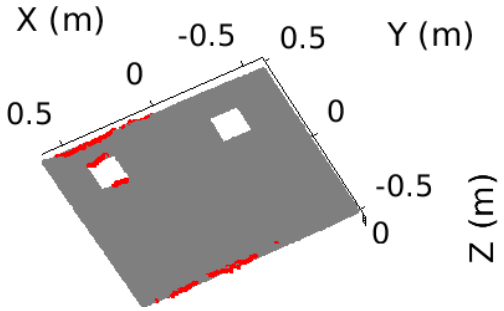

Fig. 5: Reconstructed boundary points from a single sweep along the sensor z-axis.

V. Conclusion

In this work, we have presented a novel framework for reconstructing surface points observed by an imaging sonar sensor. We have derived the 2D Fermat flow equation, which may be applied to reconstruct 3D surface points from single observations by simply translating the sensor to estimate spatial gradients. While this approach is primarily useful for reconstructing points of specular reflection in the 2D cross-sectional plane, we describe how the same sensor motion may be used to accurately reconstruct boundary points based on multiple observations. We demonstrate the effectiveness of our proposed algorithm in simulation as well as on real-world datasets collected in a test tank environment.

In future work, objects with some convex surfaces should be reconstructed with our proposed method in real-world experiments to highlight its advantage over space carving. We note that the detection of discontinuities corresponding to Fermat paths is one of the main sources of error in our real-world experiments and could be improved in future works as well.

We have treated image columns as separate sensors in this work and solve for the 3D location of points across image columns independently. Information from Fermat paths in neighboring image columns could be used as a form of regularization to improve the estimation of surface points. Additionally, Xin et al. [24] describe an optimization procedure to generate more accurate surface estimates in the presence of noise in the Fermat pathlength gradient estimates in the NLOS scenario, which could be applied to the problem of imaging sonar reconstruction.

[3]https://gdmissionsystems.com/products/underwater-vehicles/bluefin-hauv

REFERENCES

[1] P. F. Alcantarilla, J. Nuevo, and A. Bartoli, "Fast explicit diffusion for accelerated features in nonlinear scale spaces," in *British Machine Vision Conf. (BMVC)*, 2013.

[2] M. D. Aykin and S. Negahdaripour, "Forward-look 2-D sonar image formation and 3-D reconstruction," in *Proc. of the IEEE/MTS OCEANS Conf. and Exhibition*, 2013, pp. 1–10.

[3] ——, "On 3-D target reconstruction from multiple 2-D forward-scan sonar views," in *Proc. of the IEEE/MTS OCEANS Conf. and Exhibition*, May 2015, pp. 1949–1958.

[4] ——, "Modeling 2-D lens-based forward-scan sonar imagery for targets with diffuse reflectance," *IEEE J. of Oceanic Engineering*, vol. 41, no. 3, pp. 569–582, 2016.

[5] ——, "Three-dimensional target reconstruction from multiple 2-d forward-scan sonar views by space carving," *IEEE J. of Oceanic Engineering*, vol. 42, no. 3, pp. 574–589, 2016.

[6] N. Brahim, D. Gueriot, S. Daniel, and B. Solaiman, "3D reconstruction of underwater scenes using DIDSON acoustic sonar image sequences through evolutionary algorithms," in *Proc. of the IEEE/MTS OCEANS Conf. and Exhibition*, Jun. 2011.

[7] T. Guerneve and Y. Petillot, "Underwater 3D reconstruction using blueview imaging sonar," in *Proc. of the IEEE/MTS OCEANS Conf. and Exhibition*, 2015, pp. 1–7.

[8] T. Guerneve, K. Subr, and Y. Petillot, "Three-dimensional reconstruction of underwater objects using wide-aperture imaging sonar," *J. of Field Robotics*, vol. 35, no. 6, pp. 890–905, 2018.

[9] T. A. Huang and M. Kaess, "Towards acoustic structure from motion for imaging sonar," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, Oct. 2015, pp. 758–765.

[10] T. Huang and M. Kaess, "Incremental data association for acoustic structure from motion," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, Oct. 2016, pp. 1334–1341.

[11] T. Huang, "Acoustic structure from motion," Master's thesis, Carnegie Mellon University, Pittsburgh, PA, May 2016.

[12] J. Li, M. Kaess, R. Eustice, and M. Johnson-Roberson, "Pose-graph SLAM using forward-looking sonar," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 2330–2337, 2018.

[13] N. T. Mai, H. Woo, Y. Ji, Y. Tamura, A. Yamashita, and H. Asama, "3-D reconstruction of underwater object based on extended kalman filter by using acoustic camera images," *International Federation of Automatic Control PapersOnLine*, vol. 50, no. 1, pp. 1043–1049, 2017.

[14] S. Negahdaripour, "On 3-D motion estimation from feature tracks in 2-D FS sonar video," *IEEE Trans. Robotics*, vol. 29, no. 4, pp. 1016–1030, Aug. 2013.

[15] ——, "Application of forward-scan sonar stereo for 3-D scene reconstruction," *IEEE J. of Oceanic Engineering*, 2018.

[16] Y. S. Shin, Y. Lee, H. T. Choi, and A. Kim, "Bundle adjustment from sonar images and SLAM application for seafloor mapping," in *Proc. of the IEEE/MTS OCEANS Conf. and Exhibition*, Oct. 2015, pp. 1–6.

[17] C.-Y. Tsai, K. N. Kutulakos, S. G. Narasimhan, and A. C. Sankaranarayanan, "The geometry of first-returning photons for non-line-of-sight imaging," in *Proc. IEEE Intl. Conf. Computer Vision and Pattern Recognition*, 2017, pp. 7216–7224.

[18] Y. Wang, Y. Ji, H. Woo, Y. Tamura, A. Yamashita, and A. Hajime, "3D occupancy mapping framework based on acoustic camera in underwater environment," *IFAC-PapersOnLine*, vol. 51, no. 22, pp. 324–330, 2018.

[19] Y. Wang, Y. Ji, H. Woo, Y. Tamura, A. Yamashita, and H. Asama, "Three-dimensional underwater environment reconstruction with graph optimization using acoustic camera," in *IEEE/SICE Intl. Symp. on System Integration (SII)*, 2019, pp. 28–33.

[20] E. Westman and M. Kaess, "Wide aperture imaging sonar reconstruction using generative models," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2019.

[21] ——, "Degeneracy-aware imaging sonar simultaneous localization and mapping," *IEEE J. of Oceanic Engineering*, 2019.

[22] E. Westman, A. Hinduja, and M. Kaess, "Feature-based SLAM for imaging sonar with under-constrained landmarks," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, May 2018, pp. 3629–3636.

[23] E. Westman, I. Gkioulekas, and M. Kaess, "A volumetric albedo framework for 3D imaging sonar reconstruction," in *submission to IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2020.

[24] S. Xin, S. Nousias, K. N. Kutulakos, A. C. Sankaranarayanan, S. G. Narasimhan, and I. Gkioulekas, "A theory of Fermat paths for non-line-of-sight shape reconstruction," in *Proc. IEEE Intl. Conf. Computer Vision and Pattern Recognition*, 2019, pp. 6800–6809.

[25] Y. Yang and G. Huang, "Acoustic-inertial underwater navigation," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, May 2017, pp. 4927–4933.