Special Issue: Unmanned Marine Systems

Regular Article

Feature-based Bathymetric Matching of Autonomous Underwater Vehicle Transects using Robust Gaussian Processes

Øystein Sture[1] and Martin Ludvigsen[2]

[1]Department of Marine Technology, Norwegian University of Science and Technology (NTNU), Trondheim, Norway

[2]Norwegian University of Science and Technology (NTNU), Center for Autonomous Marine Operations and Systems (AMOS), Trondheim, Norway

Abstract: Autonomous underwater vehicles have become essential tools for the collection of high-resolution bathymetric and geophysical data. An inertial navigation system, aided by a Doppler velocity log and a surface-mounted ultrashort baseline acoustic positioning system, is generally capable of providing sufficient accuracy at conventional survey scales. However, the accuracy decreases with depth, and the resulting relative positioning across different transects may not be sufficient to resolve features of interest at fine scales. This work presents a method for accurate coregistration of the position of adjacent transects. The approach is based on detecting and matching local features in overlapping multibeam echosounder swaths. The navigational errors for the transects are taken to be described by latent Gaussian processes, observed through these matches. The hyperparameters of the Gaussian process are learned from the data themselves and do neither require tuning of filter parameters nor intimate knowledge of the autonomous system or its sensor configuration. The proposed method is robust to outliers by considering a non-Gaussian observation model. The approach is demonstrated on a data set collected at the Arctic Mid-Ocean Ridge (AMOR). The method can be used to construct high-resolution bathymetric models from repeated passes over the same area or to accurately coregister other sensors such as cameras, subbottom profilers, and magnetometers. The primary contribution of this work is the application of feature-based matching of bathymetry with a robust Gaussian process.

Keywords: GPS-denied operation, localization, mapping, position estimation, underwater robotics

1. Introduction

Autonomous underwater vehicles (AUVs) are untethered sensor carrying platforms capable of performing tasks without real-time interaction by an operator. AUVs have virtually replaced the

Received: 15 April 2021; revised: 2 November 2022; accepted: 29 January 2023; published: 3 April 2023.

Correspondence: Øystein Sture, Department of Marine Technology, Norwegian University of Science and Technology (NTNU), Trondheim, Norway, Email: oystein.sture@ntnu.no

DOI: https://doi.org/10.55417/fr.2023017

deep-tow survey tools previously used for deep-water petroleum-industry applications (Campbell et al., 2015). Moving the sensor carrying platform closer to the seabed increases the attainable spatial resolution of the collected data significantly. Geomorphological studies of small-scale features of interest, such as erosive processes, hydrothermal vents, and coral mounds, are, for example, not resolved in sufficient detail by ship-mounted instruments (Huvenne et al., 2018). Operating in closer proximity to the seabed also permits the use of instruments with stricter range requirements, such as optical sensors. Images collected by AUVs have, for example, been used as ground truth for high-resolution AUV bathymetry in order to provide semiquantitative assessments of manganese nodule coverage (Peukert et al., 2018). As the spatial resolution of the collected data increases, the navigational requirements for coregistering these data also becomes stricter.

It is preferable to use the instrument providing the highest spatial resolution and accuracy as a navigational aid if possible. Camera-based navigation aiding has therefore been actively researched. The use of feature detectors and structure from motion (SfM) together with simultaneous localization and mapping (SLAM) algorithms can reset the navigational errors by detecting a loop closure with previously visited areas. In practical applications, this can be carried out by planning self-intersections in order to increase the chance of loop-closure (Johnson-Roberson et al., 2010). Camera-based methods are most effective on AUVs with hovering capability, however, as they can move in closer proximity and at a lower velocity without loss of controllability (Paull et al., 2014). Survey-class AUVs are built to maximize the endurance of the vehicle and data acquisition at higher altitudes, and are generally not capable of hovering near the seafloor. Underactuated AUVs may struggle to maintain a fixed altitude setpoint in areas of rough topography, such as mid-ocean ridges. The quality of camera images varies significantly with fluctuations in altitude, due to the attenuation of light in water. For this reason, the performance of methods based around matching of images, such as visual SLAM, may be inconsistent.

This work presents a method for relative coregistration of adjacent transects based on multibeam echosounder data collected in near-seabed conditions onboard an underactuated AUV. The acoustic data have a wider swath compared to optical imagery, and sufficient overlap is therefore easier to attain. The chosen approach is feature-based, meaning that landmarks are found in each transect independently and matched using local feature descriptors. Exactly as the name suggests, this approach requires local variation in the bathymetry on a scale large enough to be resolved by the acoustic measurement but still small enough to provide unique features on a reasonable scale. Feature-based approaches for underwater localization has previously been investigated for side-scan sonars (Aulinas et al., 2010; King et al., 2017; Nguyen et al., 2012), where the intensity of the acoustic returns are used. Similarly, acoustic backscatter data from a multibeam echosounder has been used to coregister multibeam echosounder transects with side-scan sonar transects (Shang et al., 2019). For bathymetric sonar data, emphasis has been placed on featureless approaches in combination with SLAM techniques due to the difficulty in defining features recognizable from different directions (Roman and Singh, 2007). One example of this is bathymetric distributed particle SLAM (BPSLAM), where the depth soundings are registered in a grid-map and fused using an extended information filter (Barkby et al., 2011). Point cloud matching algorithms, such as the iterated closest point algorithm (ICP), can be used to relate point clouds by translations and rotations—but can converge to local minima and yield suboptimal results (Besl and McKay, 1992). However, ICP has successfully been used for a SLAM formulation, by extending it in a probabilistic setting (Palomer et al., 2016). This extension is able to account for the horizontal uncertainty being larger than the vertical and ignores featureless areas through a subsampling heuristic. Bathymetry from multibeam sonars are collected line-by-line, and these methods must consider subsets of the full transects in order to account for gradual navigational drift.

The data in this work were collected in an area on the mid-ocean ridge where the volcanic and tectonic activity provides a good basis for extracting salient bathymetric features (e.g., pillow lava formations). The methodology is not intended as a general replacement for featureless approaches, but is an attractive alternative or supplement in areas where such salient features are available.

1.1. Underwater Vehicle Navigation

Global navigation satellite systems (GNSS) are not available underwater. State of the art inertial navigation systems (INS) aided by a Doppler velocity log (DVL), which measures the velocity relative to the seabed or surrounding water body, can enable operation for extended durations. The INS typically fuses the measurements in a probabilistic state estimation filter, e.g., Kalman filter, using models of the sensor noise characteristics, kinematics, and/or rigid-body kinetics. The best DVL-INS systems can achieve an along-track drift of 0.1% of distance traveled when running in a straight line with a DVL bottom lock, which is the worst-case scenario in terms of bias estimation (Hagen et al., 2007). Ignoring installation misalignment, the horizontal drift is primarily caused by low-frequency errors of the DVL and heading error (Hegrenaes et al., 2016). The induced position error by the DVL is close to a magnitude larger than that of the accelerometers and gyroscopes for straight-line trajectories (Jalving et al., 2003). There are numerous error sources for the DVL, of which many scale with the velocity, thus worsening the performance in the main direction of travel. Examples of this are absorption bias, terrain bias, sound of speed bias, side-lobes, beam angle, and clock drift (Taudien and Bilén, 2017). Many of these are time-varying and dependent on external factors. Acoustic positioning relative to a transponder with a known global position, such as an ultrashort baseline (USBL) system mounted on a ship, may be used to bound the navigational drift over time and estimate slow-varying systematic errors. Although this may reduce the uncertainty to the order of meters, going beyond this generally requires auxiliary data sources or a bottom-mounted acoustic positioning network unless operating at shallow depths. In the short term, micro delta position aiding can provide auxiliary pseudovelocity measurements from small displacements as measured by a synthetic aperture sonar or camera (Hagen et al., 2007). Bathymetric measurements can be compared with a known map in terrain aided navigation (TAN) or with previously visited locations using SLAM (Paull et al., 2014).

1.2. Gaussian Processes

A Gaussian process is a generalization of the Gaussian probability distribution, describing random variables, to a stochastic process which describes the properties of functions (Rasmussen and Williams, 2006). The function values of a GP have a joint Gaussian distribution. A GP is completely specified by its mean function (m) and covariance kernel (κ). The mean is often taken to be zero. This does not imply that the mean of the posterior process is confined to be zero, however (i.e., when conditioned on observations).

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), \kappa(\mathbf{x}, \mathbf{x}')). \tag{1}$$

A valid covariance matrix is symmetric and positive semidefinite. The covariance of a Gaussian process is for this reason typically specified through a kernel function, which can ensure that these properties are present.

$$\mathrm{Cov}\left\{f(\boldsymbol{x}), f(\boldsymbol{x'})\right\} = \kappa(\boldsymbol{x}, \boldsymbol{x'}). \tag{2}$$

Many possible choices exist for the kernel; linear regression, splines, and Kalman filters are all examples of GPs with particular kernels (Duvenaud, 2014).

The posterior of a Gaussian process with a Gaussian likelihood function (i.e., observation model) can be exactly inferred in closed form. This incurs a computational complexity on the order of $\mathcal{O}(n^3)$ and memory requirements of $\mathcal{O}(n^2)$. Exact inference therefore quickly becomes untractable for larger datasets, and various approximations have been proposed in order to improve its computational characteristics (Liu et al., 2020). In this work, sparse variational GPs are used, where the posterior distribution of the full GP is approximated using a low-rank representation through a set of inducing inputs. The placement of these inducing inputs and kernel hyperparameters are determined through variational optimization by maximizing a lower bound to the true log marginal likelihood (Titsias, 2009). This lower bound can be optimized using stochastic variational inference (SVI), where the

variational objective is updated based on noisy gradient estimates from subsamples of the full data set (i.e., batches of observations) (Hoffman et al., 2013). The variational inference can also be efficiently used to approximate the posterior of models with non-Gaussian likelihoods (Hensman et al., 2013), which is also exploited in this work in order to be robust to outliers.

Linear Coregionalization

Originally developed in the context of geostatistics under the name cokriging, the linear model of coregionalization (LMC) is an approach to multivariate modeling (Álvarez et al., 2012). The LMC models the outputs as a linear combination of several independent latent functions. These can be modeled as Gaussian processes $\boldsymbol{g}(\boldsymbol{x}) \in \mathbb{R}^L$, which are mixed by a matrix $\boldsymbol{W} \in \mathbb{R}^{P\times L}$ to form the outputs $\boldsymbol{f}(\boldsymbol{x}) \in \mathbb{R}^P$,

$$\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{W}\boldsymbol{g}(\boldsymbol{x}). \tag{3}$$

When applied to a sparse Gaussian process, the inducing variables can be defined in the latent space, with the mean and variance projected through the mixing matrix in order to form the outputs. The mean and variance of the outputs are then given by the following equations (van der Wilk et al., 2020).

$$\boldsymbol{\mu}_f = \boldsymbol{W}\boldsymbol{\mu}_g \quad \text{and} \quad \boldsymbol{\Sigma}_f = \boldsymbol{W}\boldsymbol{\Sigma}_g\boldsymbol{W}^T. \tag{4}$$

Here, the latent covariance matrix $\boldsymbol{\Sigma}_g$ is taken to be diagonal since the latent functions are considered independent.

1.3. Preliminaries

Matrices are written in bold capital letters and vectors are written in bold lower case letters. The short-hand notation c_a and s_a is used for cosines and sines with respect to the specified angle, a. A variable in the Euclidean space with dimension n is denoted $\mathbb{R}^n$, while matrices of dimension $n \times m$ are denoted $\mathbb{R}^{n\times m}$. Superscript in curly brackets denotes the reference frame to which a given vector is expressed. For example, $p^{\{n\}}$ is a position in the north-east-down (NED) frame. The reference frames used are NED (n) and BODY (b). A rotation matrix $\boldsymbol{R}_a^b \in \mathrm{SO}(3)$ between reference frames uses a subscript for the frame transformed from, and superscript for the frame transformed to. The following notation is used for rotations about the principal axes:

$$\underset{\boldsymbol{R}_{x,\phi}}{\begin{bmatrix} 1 & 0 & 0 \\ 0 & c_\phi & -s_\phi \\ 0 & s_\phi & c_\phi \end{bmatrix}} \underset{\boldsymbol{R}_{y,\theta}}{\begin{bmatrix} c_\theta & 0 & s_\theta \\ 0 & 1 & 0 \\ -s_\theta & 0 & c_\theta \end{bmatrix}} \underset{\boldsymbol{R}_{z,\psi}}{\begin{bmatrix} c_\psi & -s_\psi & 0 \\ s_\psi & c_\psi & 0 \\ 0 & 0 & 1 \end{bmatrix}}.$$

The transformation from BODY to NED is denoted $\boldsymbol{R}_b^n(\boldsymbol{\Theta}) = \boldsymbol{R}_{z,\psi}\boldsymbol{R}_{y,\theta}\boldsymbol{R}_{x,\phi}$, where $\boldsymbol{\Theta} = [\phi, \theta, \psi]^T$ is the Euler angle parametrization following the zyx convention.

The skew-symmetric matrix $\boldsymbol{S}(\boldsymbol{\lambda}) \in \mathrm{SS}(3)$ denotes the vector cross product operator defined as follows:

$$\boldsymbol{S}(\boldsymbol{\lambda}) = \begin{bmatrix} 0 & -\lambda_3 & \lambda_2 \\ \lambda_3 & 0 & -\lambda_1 \\ -\lambda_2 & \lambda_1 & 0 \end{bmatrix}, \quad \boldsymbol{\lambda} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix}. \tag{5}$$

2. Methods

This section outlines the methods used in order to detect local features in the multibeam echosounder data, and to find possible matches between transects. These are subsequently taken to be observations of latent Gaussian processes modeling the relative error as functions correlated in time. This enables the errors to be estimated while enforcing a smoothness constraint on the possible solutions. In these transects, approximately 200 m in length, there is enough within-transect variation to make

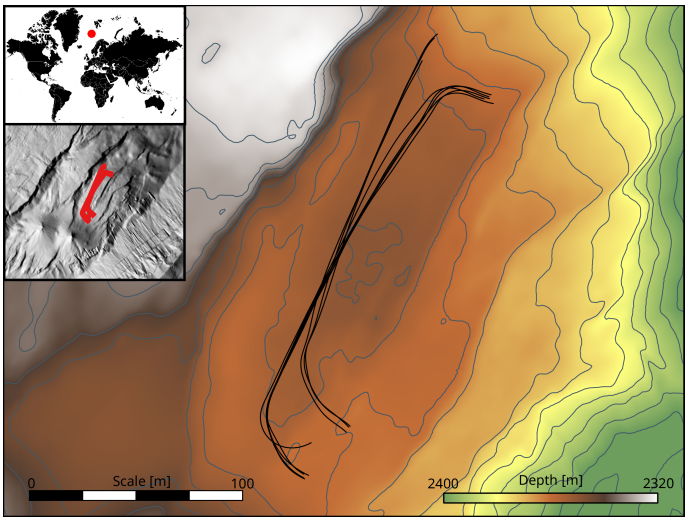


Figure 1. Operational area in which the dataset was collected, at depth of about 2350 m. The smaller map on the lower left shows an overview in which the Loki's Castle hydrothermal vent site is visible. The interval between the depth contour lines is 5 m.

a global matching impractical. While it is possible to mitigate this by subdividing into smaller patches, selecting appropriate patch sizes and enforcing smoothness or consistency between them can be challenging. Additionally, nonparametric methods such as Gaussian processes can find the error model from the data, which may otherwise be difficult to determine.

2.1. Experimental Setup and Dataset

The dataset used in this work was collected by a Kongsberg Hugin 1000 AUV during an expedition to the Atlantic Mid-Ocean ridge (Ludvigsen et al., 2016). Survey lines were planned along a shelf on the northeastern flank of the Loki's Castle hydrothermal vent site. Here, near-seabed transects were planned in alternating directions at altitudes between 6.5 to 15 m. Sending an underactuated AUV in close proximity to the seabed in an area of high bathymetric relief such as this is not without its risks. The forward facing sonar used for collision avoidance and terrain tracking was mounted in a vertical fan, and provides little predictive information while turning. The AUV was therefore instructed to gain altitude at the ends of the transects before making a turn. A map of the eight transects and surrounding bathymetry is shown in Figure 1.

The AUV was equipped with a navigation-grade Honeywell HG9900 IMU, Teledyne RDI WHN 300 kHz DVL, and was aided by a ship-mounted Kongsberg HiPAP USBL 502 USBL system. The data were collected using a Kongsberg EM2040 multibeam echosounder, with 400 beams at a center frequency of 400 kHz. This sonar is capable of of dynamic near-field focusing, in order to correct for nonoptimal beam-forming in its near-field regime, such as during near-seabed data collection (Lurton, 2002). The navigation data used in this work were first post-processed by NavLab, which performs offline smoothing based on an Error-State Kalman Filter (ESKF) and a Rauch-Tung-Striebel (RTS) smoother implementation (Gade, 2005; Willumsen and Hegrenæs, 2009). The smoothing ensures that the nominal navigation estimates are free from discontinuities due to acoustic position updates. This is important, as discontinuities violate the smoothness prior of the Gaussian process.

2.2. Multibeam Measurements

A multibeam sonar obtains depth measurements by transmitting a fan-shaped acoustic pulse, reflected by the seabed, and later measured by a transceiver sensitive to acoustic signals orthogonal

to the transmit pulse. The product of both the transmit and receive characteristics forms a directional measurement most sensitive to returns along a beam, or main lobe. Time of flight and digital beam-steering is applied in order to record a set of depth soundings along the fan-shaped transmit pulse. The measured scalar range (r), known bearing (α), and elevation (β), can be used to compute the relative displacement of a sounding in the north-east-down (NED) frame, $\boldsymbol{r}^{\{n\}}$. This can be added to the vehicle position in the same frame $\boldsymbol{p}_V^{\{n\}}$ to obtain the absolute position of the sounding $\boldsymbol{p}_S^{\{n\}}$:

$$\begin{aligned}\boldsymbol{p}_S^{\{n\}} &= \boldsymbol{p}_V^{\{n\}} + \boldsymbol{r}^{\{n\}} \\ &= \boldsymbol{p}_V^{\{n\}} + \boldsymbol{R}_b^n(\boldsymbol{\Theta}) \left[\boldsymbol{p}_M^{\{b\}} + \boldsymbol{R}_{z,\alpha}\boldsymbol{R}_{y,\beta}\left[0,0,r\right]^T\right].\end{aligned} \tag{6}$$

Here, the measured range is rotated and translated to the body-fixed vehicle frame about the z-y axes according to the bearing, elevation, and multibeam lever arm ($\boldsymbol{p}_M^{\{b\}}$). This is subsequently related to the NED frame according to the vehicle attitude (Euler angle representation, $\boldsymbol{\Theta}$). Equation 6 assumes that ray-bending due to a changing sound-velocity profile has been accounted for.

An approximate measurement covariance can be derived through a first order linear approximation. Assuming that the lever arm is known and fixed, the stochastic variables are the vehicle position, vehicle attitude, the direction of the depth sounding, and the range of the depth sounding. Zero-mean perturbations are added in the local beam frame in order to represent the uncertainty in the beam direction (e.g., beam footprint, beam-forming fluctuation):

$$\boldsymbol{p}_S^{\{n\}} = \boldsymbol{p}_V^{\{n\}} + \boldsymbol{R}_b^n(\boldsymbol{\Theta}) \left[\boldsymbol{p}_M^{\{b\}} + \boldsymbol{R}_{z,\alpha}\boldsymbol{R}_{y,\beta}\boldsymbol{R}_{y,\delta_2}\boldsymbol{R}_{x,\delta_1}\left[0,0,r\right]^T\right]. \tag{7}$$

The Jacobian, $\boldsymbol{J}_z$, for the variables $\boldsymbol{z} = [\phi, \theta, \psi, \delta_1, \delta_2, r]$ can be used to approximate the propagated uncertainty as follows:

$$\text{Cov}\left\{\boldsymbol{p}_S^{\{n\}}\right\} \approx \text{Cov}\left\{\boldsymbol{p}_V^{\{n\}}\right\} + \boldsymbol{J}_z\boldsymbol{\Sigma}_z\boldsymbol{J}_z^T. \tag{8}$$

The covariance matrix $\boldsymbol{\Sigma}_z$ is a diagonal matrix containing the variances of the variables. In this work, the variances of the Euler angles are taken from the post-processed navigation, and the variance in the range is computed from the quality factor reported by the multibeam echosounder. The uncertainty in the direction is harder to quantify, and is therefore set to the -3 dB beam-width of the echosounder. The above has previously been used to derive the approximate propagated uncertainty of an ultrashort baseline position measurement (Sture et al., 2020).

2.3. Feature Extraction and Matching

This section describes the process of computing key points and features from the bathymetric measurements. The transects are first gridded to equidistant grids, such that each pixel has a square footprint in meters. The scaling does not need to be equal across the transects. The image grids are defined by affine transformations ($\boldsymbol{A}$) relating the north-east world coordinates ($p_n^{\{n\}}$, $p_e^{\{n\}}$) to the image coordinates (u, v). This relationship is shown below for a single sonar point:

$$\begin{bmatrix}u\\v\end{bmatrix} = \boldsymbol{A}\begin{bmatrix}p_n^{\{n\}}\\p_e^{\{n\}}\\1\end{bmatrix} = \begin{bmatrix}ac_\theta & -as_\theta & t_n\\as_\theta & ac_\theta & t_e\end{bmatrix}\begin{bmatrix}p_n^{\{n\}}\\p_e^{\{n\}}\\1\end{bmatrix}. \tag{9}$$

Here, t_n and t_e are translations along the x-y axes, a is a scaling factor, and θ the orientation. The affine transformation is found by minimizing the reprojection error using RANSAC followed by a Levenberg-Marquardt refinement step, implemented in the OpenCV library (Bradski, 2000). This is done in order to obtain an image grid which is aligned with the primary survey direction. It is also

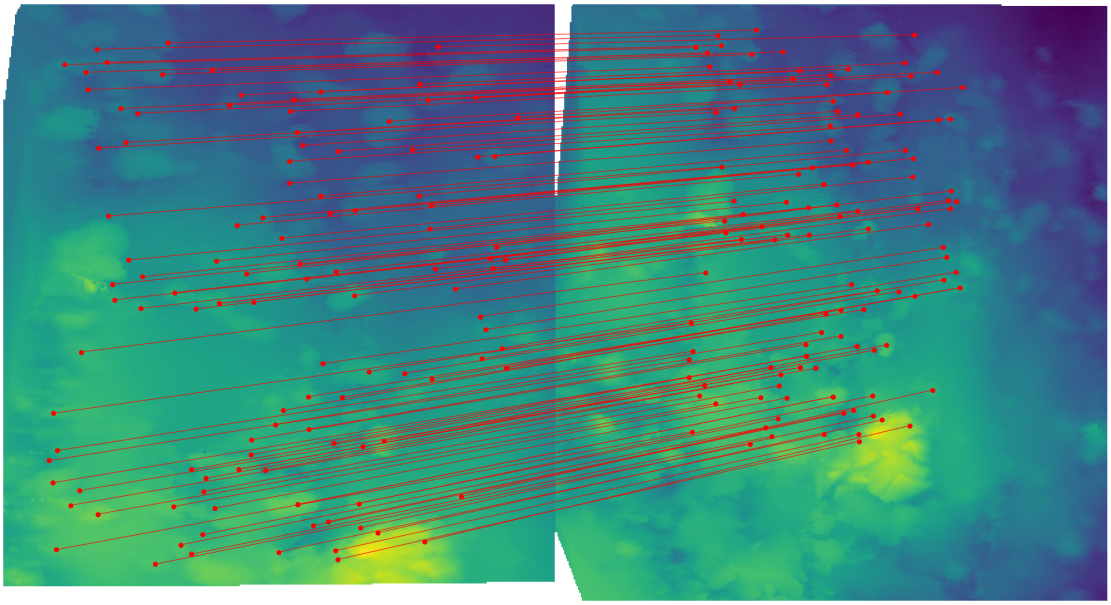


Figure 2. Results from feature-based matching on a subset of two transects over the same area at different altitudes. The images were constructed through Delaunay triangulation and linear interpolation to the grid defined by the affine transformation in Equation 9.

possible to simply use a north-east oriented grid with sufficiently high resolution, but will consist of larger areas with no data compared to the oriented grid. Once the affine transformation has been defined, the depth values are linearly interpolated to the grid from the scattered multibeam points using a Delaunay triangulation.

Scale-invariant feature transform (SIFT) is used as the feature descriptor in this work. Other image-based feature descriptors, such as ORB and BRISK were attempted, but had a tendency to cluster around areas with sharp edges, due to for example interpolation errors or acoustic artifacts. SIFT uses gradient magnitudes and orientations computed for windows surrounding the key point as its feature descriptor. The SIFT key points were matched using a k-nearest neighbor search in the SIFT feature space. A threshold of 0.8 on the ratio between the feature distance (L_1 norm) to the closest and second closest neighbor was used to accept matches (Lowe, 2004). Figure 2 shows an example of matches between two different transects. Note that the scale differs between the transects, but is correctly detected by SIFT due to its scale invariance. Scale invariance is not strictly necessary, as the true scale of the measurements is known, but enables reusing the same features against all transects without oversampling to the same resolution. The feature descriptor should also be somewhat robust to small affine distortions, in order to account for a biased body-fixed velocity. The SIFT algorithm attains its invariance to rotations by relating the descriptor orientations to the dominant orientation of local image gradients. The original SIFT algorithm used 36 bins, each covering a 10° sector, to determine the dominant orientation. Since the data were collected by an AUV equipped with a north-seeking gyroscope, with an estimated standard deviation of 0.005° from NavLab, the orientation of the data is fairly certain. For this reason, the number of mismatches was reduced by relating the local image descriptors to the same heading θ (i.e., north).

2.4. Navigation Residual Error Estimation

The goal in this section is to take the matches found and estimate the time-variable errors in position between adjacent transects. The post-processed navigation solution is used as a prior navigation

estimate, and refined using the multibeam matches as auxiliary observations. First, an expression for a multibeam measurement in terms of a true position and a position error will be derived. This is subsequently used to derive the measurement equation for a match between two transects. The following assumes that the primary error sources are the position of the vehicle and heading error.

Continuing from Equation 6, the error states can be added as follows. The frame notation is omitted from this point for brevity, as all variables are in the north-east-down frame. The variables $\delta \boldsymbol{p}_V$ and $\boldsymbol{R}_{z,\delta\psi}$ represent the error induced by position and heading offsets, respectively:

$$\boldsymbol{p}_S + \delta \boldsymbol{p}_S = \boldsymbol{p}_V + \delta \boldsymbol{p}_V + \boldsymbol{R}_{z,\delta\psi}\boldsymbol{r}. \tag{10}$$

If we assume that the deviations in heading are modest, we can apply the following small-angle approximation in order to maintain a linear expression. For sine and cosine, this exceeds a 1% relative angular error at about 8° and 14°, respectively:

$$\boldsymbol{R}_{z,\delta\psi} = \begin{bmatrix} c_{\delta\psi} & -s_{\delta\psi} & 0 \\ s_{\delta\psi} & c_{\delta\psi} & 0 \\ 0 & 0 & 1 \end{bmatrix} \approx \begin{bmatrix} 1 & -\delta\psi & 0 \\ \delta\psi & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \boldsymbol{I}_3 + \boldsymbol{S}\left([0, 0, \delta\psi]^T\right). \tag{11}$$

The expressions for the true position and position error can be split up by separating the small-angle approximation into diagonal and skew-symmetric parts. The matrix $\boldsymbol{S}(\boldsymbol{\lambda})$ was defined in Equation 5.

$$\boldsymbol{p} + \delta\boldsymbol{p} = \overbrace{\boldsymbol{p}_V + \boldsymbol{r}}^{\boldsymbol{p}} + \overbrace{\delta\boldsymbol{p}_V + \boldsymbol{S}\left([0, 0, \delta\psi]^T\right)\boldsymbol{r}}^{\delta\boldsymbol{p}}. \tag{12}$$

The goal is to estimate the position error, $\delta\boldsymbol{p}$, based on matches between key points from the multibeam measurements. We can observe these errors indirectly by taking the difference between the estimated position for key point i and j.

$$\boldsymbol{z}_{i,j} = (\boldsymbol{p}_i + \delta\boldsymbol{p}_i) - (\boldsymbol{p}_j + \delta\boldsymbol{p}_j). \tag{13}$$

Since the true position of the depth measurement is free of errors, they are equal $\boldsymbol{p}_i = \boldsymbol{p}_j$. This leads to the following observation of the residual errors:

$$\boldsymbol{z}_{i,j} = \delta\boldsymbol{p}_i - \delta\boldsymbol{p}_j = \left(\delta\boldsymbol{p}_{V,i} + \begin{bmatrix} -\delta\psi_i r_{e,i} \\ \delta\psi_i r_{n,i} \\ 0 \end{bmatrix}\right) - \left(\delta\boldsymbol{p}_{V,j} + \begin{bmatrix} -\delta\psi_j r_{e,j} \\ \delta\psi_j r_{n,j} \\ 0 \end{bmatrix}\right). \tag{14}$$

Here, the range-measurement vectors, $\boldsymbol{r} = [r_n, r_e, r_d]^T$, are expanded into their scalar components. If the error states are assembled in a vector, the measurement can be described as a linear combination of errors in position and heading. The error in depth is not modeled, as it has a negligible drift within the same dive and provides no valuable information about the other states.

$$\boldsymbol{\delta x} = [\delta p_{V,n,i},\ \delta p_{V,n,j},\ \delta p_{V,e,i},\ \delta p_{V,e,j},\ \delta\psi_i,\ \delta\psi_j]^T, \tag{15}$$

$$\boldsymbol{W}_{i,j} = \begin{bmatrix} 1 & -1 & 0 & 0 & -r_{e,i} & r_{e,j} \\ 0 & 0 & 1 & -1 & r_{n,i} & -r_{n,j} \end{bmatrix}, \tag{16}$$

$$\boldsymbol{z}_{i,j} = \boldsymbol{W}_{i,j}\boldsymbol{\delta x}. \tag{17}$$

The mixing matrix $\boldsymbol{W}_{i,j}$ relates two matches, each with three underlying latent variables, to two observations; the deviation in north and east between the matches. If more than two transects observe the same key point, they can also be added to the mixing matrix. When a match is established from a key point to multiple transects, matches between all those transects can also

be inferred. In other words, if transect A has a match with transect B and C individually, this also means that a match between B and C can be inferred. In the case that a match already exists between B and C, the Mahalanobis distance of the propagated uncertainty from Equation 8 is used to select the most likely match. For the case of two observations per pair of transects and three latent states per transect, the mixing matrix has the following dimensions with n_t denoting the number of transects.

$$\boldsymbol{W} \in \mathbb{R}^{P \times L}, \quad L = 3n_t, \quad P = \frac{n_t!}{(n_t - 2)!}. \tag{18}$$

The objective is to estimate the latent error processes as functions of time per transect, while accounting for the presence of measurement noise and possible outliers. If the latent functions are sufficiently smooth, a Gaussian process prior can be applied. Due to the small-angle approximation, the output processes can be modeled by linear coregionalization with the mixing matrix above.

The covariance of the latent functions is described by a kernel function, which must be selected. The actual movement of the vehicle is assumed to have continuous derivatives in both position and velocity due to inertia. As noted in Section 1.1, the horizontal drift of a DVL-aided INS is primarily determined by time-varying errors in the body fixed velocity measurements. The assumption is therefore made that the residual errors after the Kalman smoothing are also dominated by biases in the body-fixed velocity. For this reason, the Matérn 5/2 kernel is selected, as a Gaussian process prior with this kernel ensures that the function is at least twice differentiable (Stein, 1999). Intuitively, this kernel strikes a good balance in smoothness of the prior, but other stationary isometric kernels were also tested with similar results. The Matérn 5/2 kernel is given by the following, where d is the Euclidean distance metric, l is a length-scale parameter scaling the input distance, and σ is the output variance scale parameter (Rasmussen and Williams, 2006). In this work, the kernel is applied to a scalar difference of time within each transect, $d = |\Delta t|$.

$$\kappa_{5/2}(d) = \sigma^2 \left(1 + \frac{\sqrt{5}d}{l} + \frac{5d^2}{3l^2}\right) \exp\left(-\frac{\sqrt{5}d}{l}\right). \tag{19}$$

In order to account for outliers in the observations, for example, due to mismatches, a Student-T likelihood function is used. This is a heavy-tailed distribution, which describes values far away from the mean as more likely compared to a Gaussian likelihood. It's possible to first fit the observations to a Gaussian process with a Student-T likelihood, reject outliers, and subsequently fit the inliers to a Gaussian process with a Gaussian likelihood function (Martinez-Cantin et al., 2018). This provides both robustness to outliers and a correct likelihood under the assumption that the inliers follows a Gaussian distribution.

The scale parameter of the likelihood, the variance parameter of the Matern kernel, and placement of the inducing inputs are optimized through stochastic gradient descent using the Adam method (Kingma and Ba, 2015). The above was built on top on GPFlow, a framework for Gaussian process inference in Tensorflow (Matthews et al., 2017; van der Wilk et al., 2020).

It is worth noting that the latent states could be modeled as one long sequence rather than dividing into distinct transects. Splitting into distinct transects discards any correlation in time between them. In other words, the end of the previous transect and start of next transect are considered to be uncorrelated. There is a computational benefit in doing this, as each transect is defined by its own set of inducing inputs rather than sharing an increasing number of inducing inputs in order to support a larger input domain. Strategies for handling boundary conditions between the partitions, such as adding pseudo-observations at the boundaries as an approximate continuity condition can be considered in the future (Park and Apley, 2018).

3. Results

Eight distinct transects were processed using the described method. The latent processes estimated for each transect are shown in Figure 3. These show the deviation in north, deviation in east,

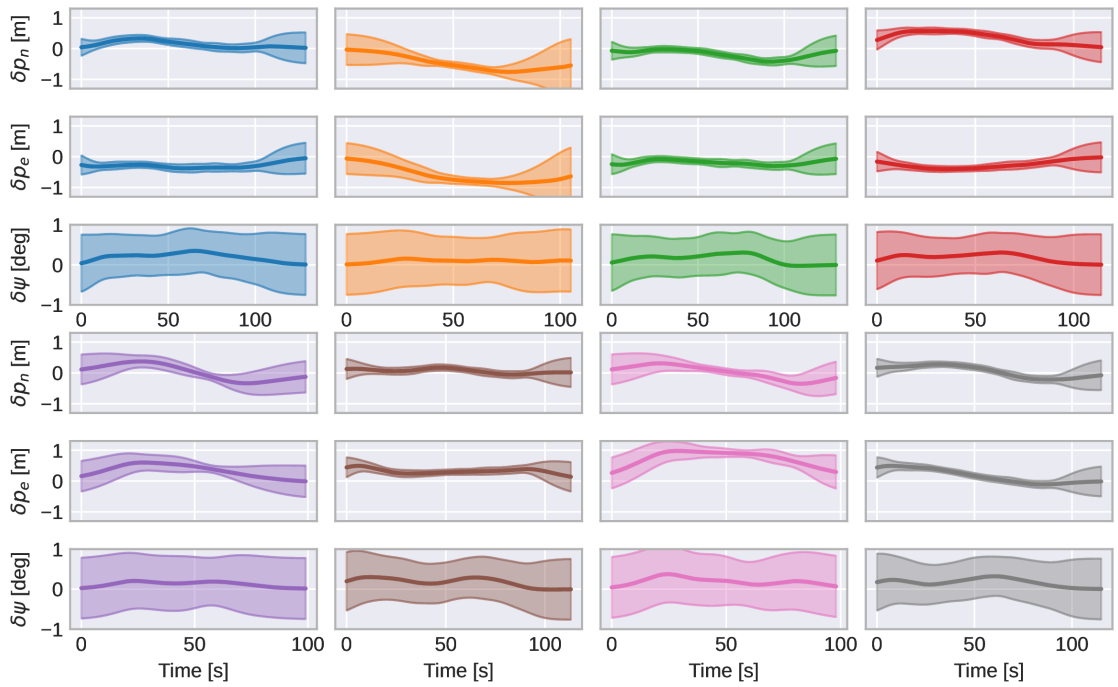


Figure 3. The estimated latent processes ($\delta p_{V,n}$, $\delta p_{V,e}$, $\delta\psi$) for each trajectory. The shaded regions are the 97.5th percentiles of the processes (95% confidence interval). The x axes are the times from start to finish for each transect and does not correspond to the same points in time.

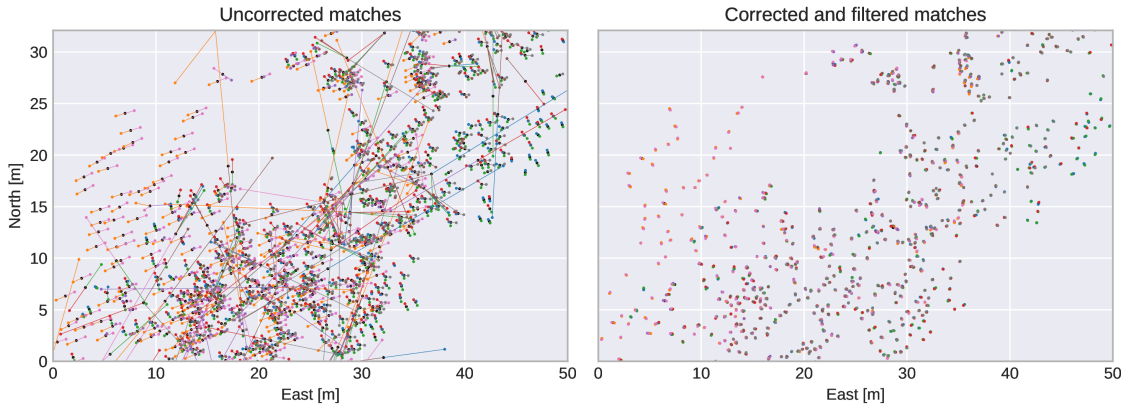


Figure 4. Comparison of the match locations from a spatial subset of the transects. Matches are plotted as points colored according to the transect, with a line connecting them to the mean of that cluster of matches (in black). The plot on the right contains the same matches, where the outliers have been removed with the Student-t likelihood, and the inliers subsequently fitted to a Gaussian process with a Gaussian likelihood function.

and heading deviation plotted against the relative time since the beginning of each transect. The shaded areas are the 97.5th percentile of the latent processes. The percentiles increase towards the start and end of the transects, a natural consequence of there being fewer observations there. This can be used to determine how trustworthy the estimates are. The error in position attributed to heading deviation is modest, which is expected given the accuracy of the ring-laser gyroscope. Figure 4 depicts key points and matches for a spatial subset of the full transects. The key points are connected to the mean of the matches by a line. The left figure contains the matches prior to

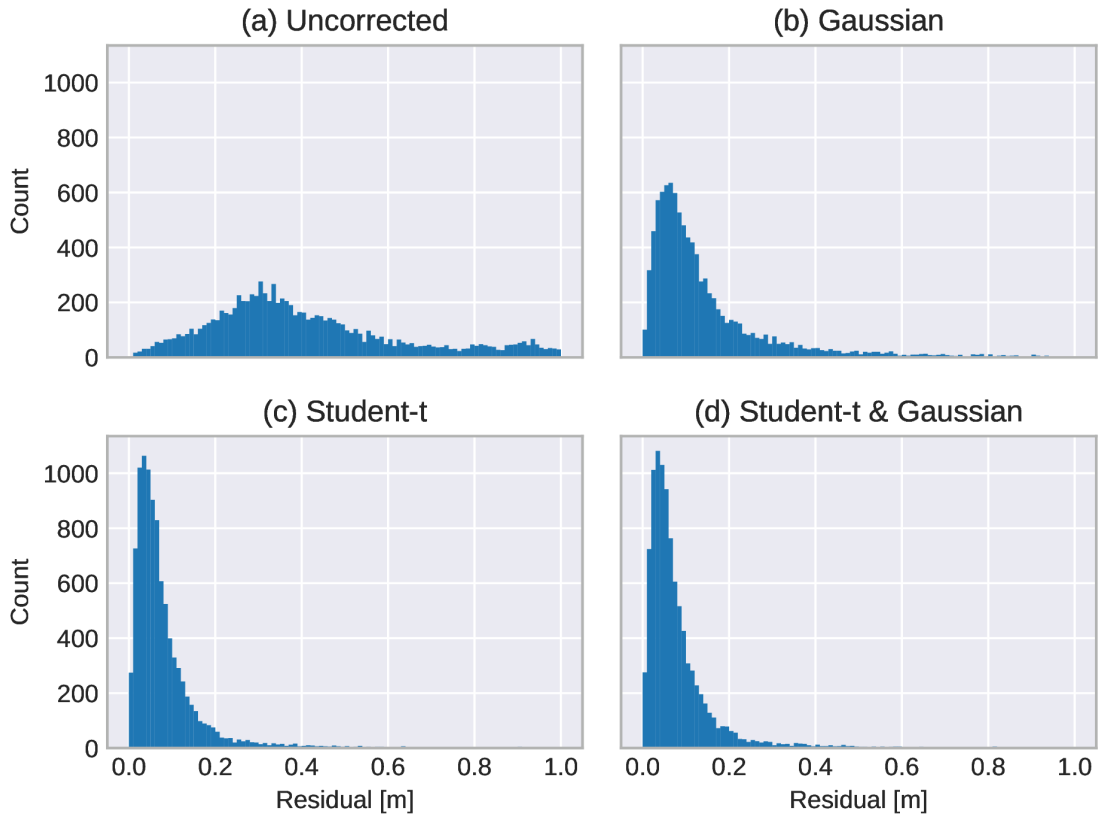


Figure 5. Histograms of residuals with respect to the mean in each cluster of matches, after subtracting the estimated error processes, for residuals below 1 m. Four sets of residuals are compared: (a) uncorrected, (b) GP with a Gaussian likelihood, (c) GP Student-t with likelihood, and (d) GP with Student-t likelihood for outlier rejection followed by a GP with a Gaussian likelihood.

any corrections, and the right figure contains corrected matches. The matches are corrected by first rejecting outliers, and then fitting a Gaussian process to the inliers.

To investigate the effect of a Gaussian process with the Student-t likelihood, the residual distances to the mean of each cluster of matches are compared after subtracting the errors estimated by the Gaussian process. Histograms for these residual errors below 1 m are shown in Figure 5. Four cases are compared: (a) uncorrected observations, (b) GP with Gaussian likelihood, (c) GP with Student-t likelihood, and (d) GP with Gaussian likelihood on inliers, after determining outliers using the result from (c). The GP with the Gaussian likelihood is able to reduce the residuals significantly, but is clearly affected by outliers when compared to the correction with a Student-t likelihood. The mean absolute error compared for only the matches defined as inliers yields the following results for the four cases; $e_a = 42.9\,\text{cm}$, $e_b = 13.8\,\text{cm}$, $e_c = 6.63\,\text{cm}$, and $e_d = 6.65\,\text{cm}$. The residuals after correcting Gaussian Student-t likelihood with outliers are nearly identical to a Gaussian likelihood with outliers removed. Observations exceeding a threshold placed on the 97.5th percentile were taken to be outliers. It would seem that the approach outlined by Martinez-Cantin et al. (2018) does not reduce the residuals in this case. It is possible that this may be caused by some mild outliers that remain.

Figure 6 and Figure 7 shows the absolute errors along the z axis between a reference transect and two other transects, respectively. A clear improvement in the overall alignment is demonstrated. The error visible in the rightmost part of Figure 6 is due to a constant offset in depth between the transects, possibly due to the sharp turn in the reference transect. The local features themselves

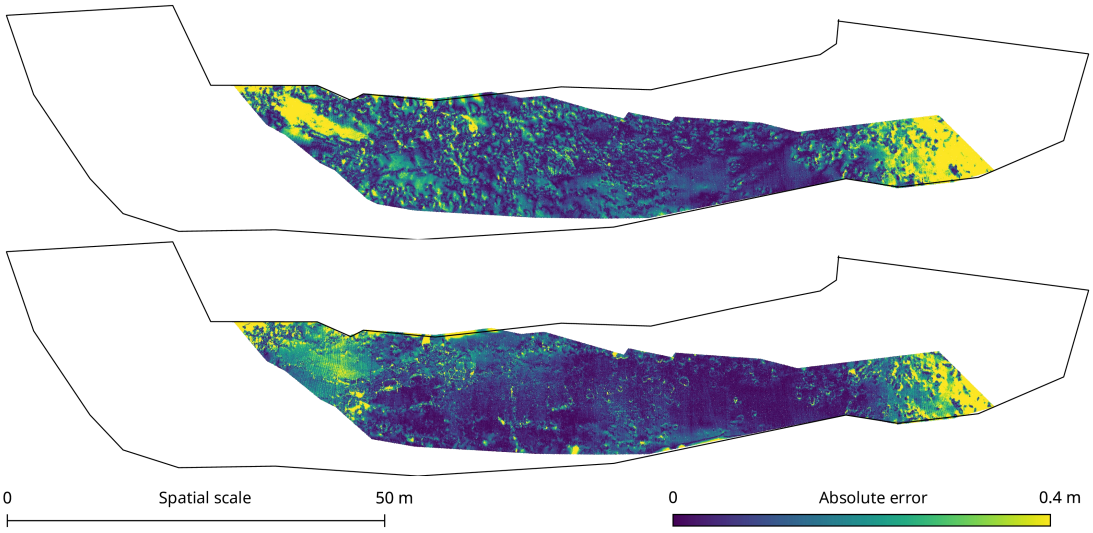


Figure 6. Absolute error in z values between a reference transect (3) and transect 5. The top and bottom figures are the errors before and after correction, respectively. The outline is the extent of the data from the reference trajectory. These transects pass in opposite directions and across.

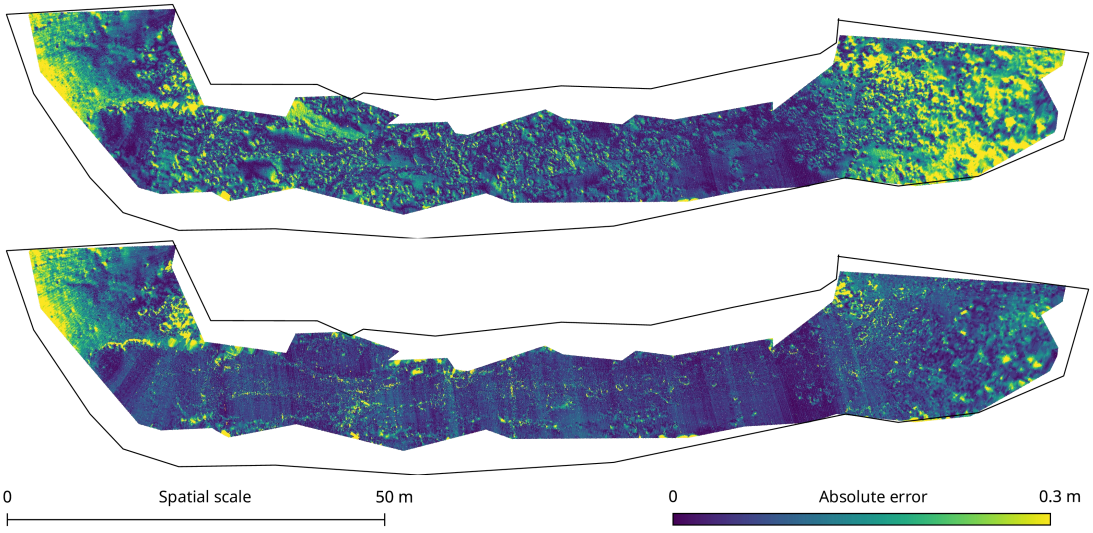
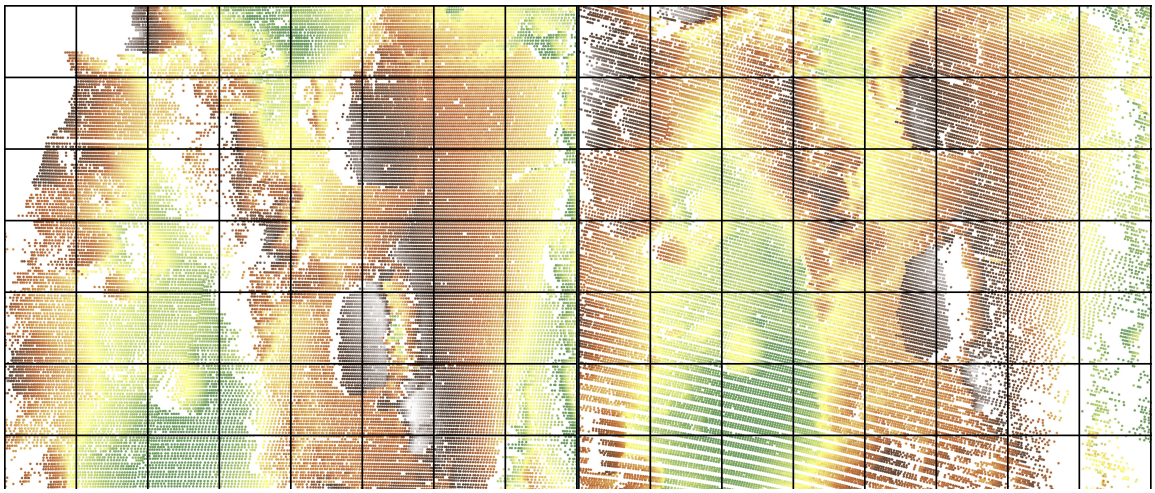


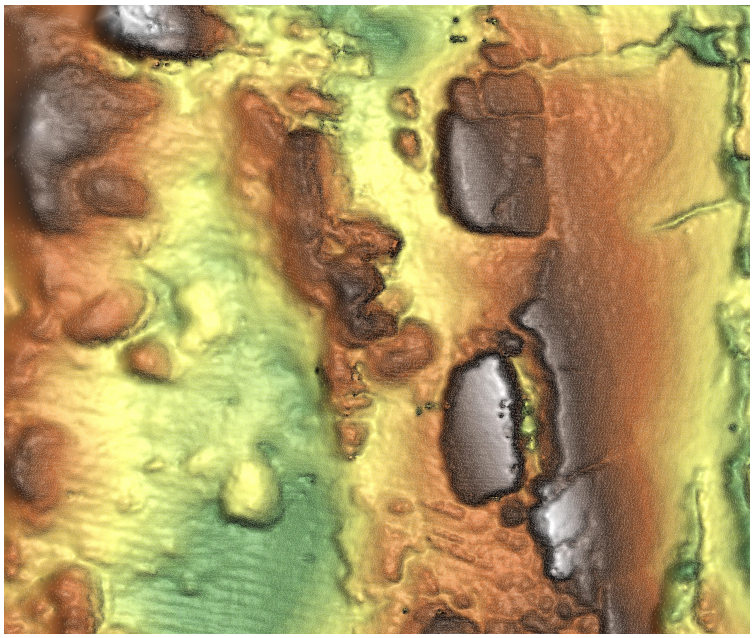
Figure 7. Absolute error in z values between a reference transect (3) and transect 6. The top and bottom figures are the errors before and after correction, respectively. The outline is the extent of the data from the reference trajectory. These transects follow the same path at different altitudes.

seem to be well matched, although the reference transect has lower resolution in the outer part of the turn. In Figure 7, small errors due to either heave or pitch are plainly visible as stripes spanning the entire swath. The leftmost part contains some constant offset error in the reference transect, possibly due to the turn similar to the previous figure.

Finally, a combination of corrected data for all eight transects in the same area is provided as a final illustration in Figure 8b. The result is a high-density point cloud, where slight shading has been applied to accentuate the features. No smoothing has been performed on the depth data themselves, and for this reason, the gradient estimates are in some places poor, which are visible as apparent holes.



(a) Two uncorrected transects.



(b) Mesh of merged data for eight corrected transects for a subset of the full transects. The combined data consist of approximately 500 000 points. No smoothing has been applied, but slight shading has been added to accentuate smaller features.

Figure 8. Comparison of multibeam echosounder point clouds for uncorrected and corrected transects in a selected area. The extent of the area is approximately 18.5 by 15.5 meters.

3.1. Future Work

The presented method relies on local features, where the majority are inliers, in order to resolve the navigational offset between the transects. Although the Student-T process makes the method robust to outliers, the quality of the results depend on the features extracted. A feature extraction method capable of operating directly on the soundings, or point cloud, rather than interpolating to a regular grid is desirable. In this work, transects with high-density soundings and few holes were investigated, and thus an approach based on conventional image processing was possible.

Given irregularly spaced data, more holes, or larger holes, other means of detecting salient features needs to be investigated. Ideally, this method should directly be capable of accounting for the sounding uncertainty. SIFT-like shape descriptors has been generalized to point clouds (Jiang et al., 2018). Feature-based matching of lidar scans for self-driving vehicles has also been investigated, for example by making neural networks learn which features are robust for matching (Lu et al., 2019).

The latent error residuals were estimated without considering the effect of the velocity. It may be possible to include prior knowledge regarding the rate of change of the latent process, since the standard deviation of the velocity measurements is known to vary according to the velocity. A strategy may also involve using the outlined method to reject outliers, and then use the matches in a conventional extended Kalman filter or error state filter formulation, where all observations are assumed to be inliers.

4. Conclusion

A method for aligning multiple autonomous underwater vehicle transects has been presented. The approach is data-driven, and does not require intimate knowledge of the sensors and target system, and finds most parameters through optimization. The approach is robust to outliers by rejecting unlikely matches based on a Student-T process. The threshold at which to reject matches depends on the quality of the collected data, and is specified by the user. The method is validated on a dataset containing varying altitudes, alternating directions, and orientations and has been found to significantly improve the fine-alignment of the multibeam echosounder data.

ORCID

Øystein Sture https://orcid.org/0000-0003-4892-2883
Martin Ludvigsen https://orcid.org/0000-0002-8553-9245

References

Álvarez, M. A., Rosasco, L., and Lawrence, N. D. (2012). Kernels for Vector-Valued Functions: A Review. *Foundations and Trends in Machine Learning*, 4(3):195–266. https://doi.org/10.1561/2200000036.

Aulinas, J., Llado, X., Salvi, J., and Petillot, Y. R. (2010). Feature based slam using side-scan salient objects. In *OCEANS 2010 MTS/IEEE SEATTLE*, Seattle, WA. https://doi.org/10.1109/OCEANS.2010.5664461.

Barkby, S., Williams, S. B., Pizarro, O., and Jakuba, M. V. (2011). A featureless approach to efficient bathymetric SLAM using distributed particle mapping. *Journal of Field Robotics*, 28(1):19–39. https://doi.org/10.1002/rob.20382.

Besl, P. J. and McKay, N. D. (1992). Method for registration of 3-D shapes. In *Sensor Fusion IV: Control Paradigms and Data Structures*, volume 1611, pages 586–606. International Society for Optics and Photonics.

Bradski, G. (2000). The OpenCV library. *Dr. Dobb's Journal of Software Tools*.

Campbell, K. J., Kinnear, S., and Thame, A. (2015). AUV technology for seabed characterization and geohazards assessment. *The Leading Edge*, 34(2):140–260. https://doi.org/10.1190/tle34020170.1.

Duvenaud, D. K. (2014). *Automatic Model Construction with Gaussian Processes (Doctoral Thesis)*. PhD thesis, University of Cambridge. https://doi.org/10.17863/CAM.14087.

Gade, K. (2005). NavLab, a Generic Simulation and Post-processing Tool for Navigation. *Modeling, Identification and Control: A Norwegian Research Bulletin*, 26(3):135–150. https://doi.org/10.4173/mic.2005.3.2.

Hagen, P. E., Hegrenæs, Ø., Jalving, B., Midtgaard, Ø., Wiig, M. S., and Hagen, O. K. (2007). Making AUVs truly autonomous. In *OCEANS 2007*. IEEE. https://doi.org/10.1109/OCEANS.2007.4449405.

Hegrenaes, Ø., Ramstad, A., Pedersen, T., and Velasco, D. (2016). Validation of a new generation DVL for underwater vehicle navigation. In *2016 IEEE/OES Autonomous Underwater Vehicles (AUV)*, pages 342–348, Tokyo, Japan. IEEE. https://doi.org/10.1109/AUV.2016.7778694.

Hensman, J., Fusi, N., and Lawrence, N. D. (2013). Gaussian processes for big data. In *Proceedings of the Twenty-Ninth Conference on Uncertainty in Artificial Intelligence*, UAI'13, pages 282–290, Arlington, Virginia, USA. AUAI Press.

Hoffman, M. D., Blei, D. M., Wang, C., and Paisley, J. (2013). Stochastic Variational Inference. *Journal of Machine Learning Research*, 14(1):1303–1347.

Huvenne, V. A., Robert, K., Marsh, L., Iacono, C. L., Le Bas, T., and Wynn, R. B. (2018). Rovs and auvs. In *Submarine Geomorphology*, Springer Geology, pages 93–108. Springer, Cham. https://doi.org/10.1007/978-3-319-57852-1_7.

Jalving, B., Bovio, E., and Gade, K. (2003). Integrated inertial navigation systems for AUVs for REA applications. In *MREP*, NATO SACLANT Undersea Research Centre, Italy.

Jiang, M., Wu, Y., Zhao, T., Zhao, Z., and Lu, C. (2018). PointSIFT: A SIFT-like Network Module for 3D Point Cloud Semantic Segmentation. *arXiv:1807.00652*.

Johnson-Roberson, M., Pizarro, O., Williams, S. B., and Mahon, I. (2010). Generation and visualization of large-scale three-dimensional reconstructions from underwater robotic surveys. *J. Field Robotics*, 27(1):21–51. https://doi.org/10.1002/rob.20324.

King, P., Anstey, B., and Vardy, A. (2017). Sonar Image Registration for Localization of an Underwater Vehicle. *The Journal of Ocean Technology*, 12(3):68–90.

Kingma, D. P. and Ba, J. (2015). Adam: A Method for Stochastic Optimization. In *International Conference for Learning Representations*, San Diego, CA, USA.

Liu, H., Ong, Y.-S., Shen, X., and Cai, J. (2020). When Gaussian Process Meets Big Data: A Review of Scalable GPs. *IEEE Trans. Neural Netw. Learning Syst.*, 31(11):4405–4423. https://doi.org/10.1109/TNNLS.2019.2957109.

Lowe, D. G. (2004). Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2):91–110. https://doi.org/10.1023/B:VISI.0000029664.99615.94.

Lu, W., Wan, G., Zhou, Y., Fu, X., Yuan, P., and Song, S. (2019). DeepVCP: An End-to-End Deep Neural Network for Point Cloud Registration. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 12–21, Seoul, Korea (South). IEEE. https://doi.org/10.1109/ICCV.2019.00010.

Ludvigsen, M., Aasly, K., Ellefmo, S. L., Hilário, A., Ramirez-Llodra, E., Søreide, F. X., Falcon-Suarez, I., Juliani, C. J., Kieswetter, A., Lim, A., et al. (2016). MarMine cruise report-Arctic Mid-Ocean Ridge 15.08. 2016-05.09. 2016. Technical Report 2016-01, Norwegian University of Science and Technology, Trondheim, Norway.

Lurton, X. (2002). *An Introduction to Underwater Acoustics: Principles and Applications*. Springer Science & Business Media.

Martinez-Cantin, R., Tee, K., and McCourt, M. (2018). Practical Bayesian optimization in the presence of outliers. In *International Conference on Artificial Intelligence and Statistics*, pages 1722–1731. PMLR.

Matthews, D. G., Alexander, G., Van Der Wilk, M., Nickson, T., Fujii, K., Boukouvalas, A., León-Villagrá, P., Ghahramani, Z., and Hensman, J. (2017). GPflow: A Gaussian process library using TensorFlow. *The Journal of Machine Learning Research*, 18(1):1299–1304.

Nguyen, H.-G., Fablet, R., Ehrhold, A., and Boucher, J.-M. (2012). Keypoint-Based Analysis of Sonar Images: Application to Seabed Recognition. *IEEE Trans. Geosci. Remote Sensing*, 50(4):1171–1184. https://doi.org/10.1109/TGRS.2011.2165848.

Palomer, A., Ridao, P., and Ribas, D. (2016). Multibeam 3D Underwater SLAM with Probabilistic Registration. *Sensors*, 16(4):560. https://doi.org/10.3390/s16040560.

Park, C. and Apley, D. (2018). Patchwork Kriging for Large-scale Gaussian Process Regression. *The Journal of Machine Learning Research*, 19(1):269–311.

Paull, L., Saeedi, S., Seto, M., and Li, H. (2014). AUV Navigation and Localization: A Review. *IEEE Journal of Oceanic Engineering*, 39(1):131–149. https://doi.org/10.1109/JOE.2013.2278891.

Peukert, A., Schoening, T., Alevizos, E., Köser, K., Kwasnitschka, T., and Greinert, J. (2018). Understanding Mn-nodule distribution and evaluation of related deep-sea mining impacts using AUV-based hydroacoustic and optical data. *Biogeosciences*, 15(8):2525–2549. https://doi.org/10.5194/bg-15-2525-2018.

Rasmussen, C. E. and Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning*. Adaptive Computation and Machine Learning. MIT Press, Cambridge, Mass.

Roman, C. and Singh, H. (2007). A Self-Consistent Bathymetric Mapping Algorithm. *J. Field Robotics*, 24(1-2):23–50. https://doi.org/10.1002/rob.20164.

Shang, Zhao, and Zhang (2019). Obtaining High-Resolution Seabed Topography and Surface Details by Co-Registration of Side-Scan Sonar and Multibeam Echo Sounder Images. *Remote Sensing*, 11(12):1496. https://doi.org/10.3390/rs11121496.

Stein, M. L. (1999). *Interpolation of Spatial Data*. Springer Series in Statistics. Springer New York, New York, NY. https://doi.org/10.1007/978-1-4612-1494-6.

Sture, Ø., Norgren, P., and Ludvigsen, M. (2020). Trajectory Planning for Navigation Aiding of Autonomous Underwater Vehicles. *IEEE Access*, pages 1–1. https://doi.org/10.1109/ACCESS.2020.3004439.

Taudien, J. Y. and Bilén, S. G. (2017). Quantifying long-term accuracy of sonar Doppler velocity logs. *IEEE Journal of Oceanic Engineering*, 43(3):764–776. https://doi.org/10.1109/JOE.2017.2735558.

Titsias, M. K. (2009). Variational Learning of Inducing Variables in Sparse Gaussian Processes. In *International Conference on Artificial Intelligence and Statistics*, volume 5 of *Proceedings of Machine Learning Research*, pages 567–574, Hilton Clearwater Beach Resort, Clearwater Beach, Florida USA. PMLR.

van der Wilk, M., Dutordoir, V., John, S. T., Artemev, A., Adam, V., and Hensman, J. (2020). A Framework for Interdomain and Multioutput Gaussian Processes. *arXiv:2003.01115*.

Willumsen, A. B. and Hegrenæs, Ø. (2009). The joys of smoothing. In *OCEANS 2009-EUROPE*, Bremen, Germany. IEEE. https://doi.org/10.1109/OCEANSE.2009.5278142.

How to cite this article: Sture, Ø., & Ludvigsen, M. (2023). Feature-based bathymetric matching of autonomous underwater vehicle transects using robust Gaussian processes. *Field Robotics, 3*, 544–559.