

# Incremental Data Association for Acoustic Structure from Motion

Tiffany A. Huang and Michael Kaess

**Abstract**—We provide a novel incremental data association method to complement our previous work on acoustic structure from motion (ASFM), which recovers 3D scene structure from multiple 2D sonar images, while at the same time localizing the sonar. Given point features extracted from multiple overlapping sonar images, our algorithm automatically finds the correspondences between the features. Our data association method uses information about the geometric correlations of the entire set of landmarks to reject spurious measurements or false positives that might otherwise have been accepted. For each new sonar measurement, the algorithm uses a gating procedure to narrow the landmark match search space. Using the pruned surviving candidate correspondences, we identify the correct hypothesis based on a posterior compatibility cost, penalizing for null matches to avoid all measurements being declared new landmarks. Unlike other methods, ASFM does not require any planar scene assumptions and uses constraints from more than two images to increase accuracy in both mapping and localization. We evaluate our algorithm in simulation and demonstrate successful data association results on real sonar images.

## I. INTRODUCTION

Mapping and state estimation have been widely explored for autonomous vehicles that operate on land and in the air. However, for an environment that spans the majority of our planet Earth, surprisingly little progress has been made towards the same autonomous abilities underwater. Autonomous underwater vehicles (AUVs) open the door to exciting new possibilities for underwater exploration such as venturing into areas too dangerous for human divers or exploring large areas much faster and more efficiently. Furthermore, AUVs have the potential to eliminate the tedium and high costs of remotely operated vehicle (ROV) missions.

In this paper, we discuss data association for simultaneous localization and mapping (SLAM) underwater. One particular challenge towards SLAM for AUVs is the necessity of non-conventional sensors such as sonar. Due to the turbidity of some water environments as well as the short propagation range of light in water, more common and well-studied sensors such as cameras and LIDAR do not work well underwater. As for localization, GPS cannot be used since radio waves are rapidly absorbed by water. In our previous work [6], we proposed a solution called acoustic structure from motion (ASFM), which uses multiple, general sonar viewpoints of the same scene to reconstruct the 3D structure of select point features while minimizing the effects of accumulating error. ASFM provides several advantages over other

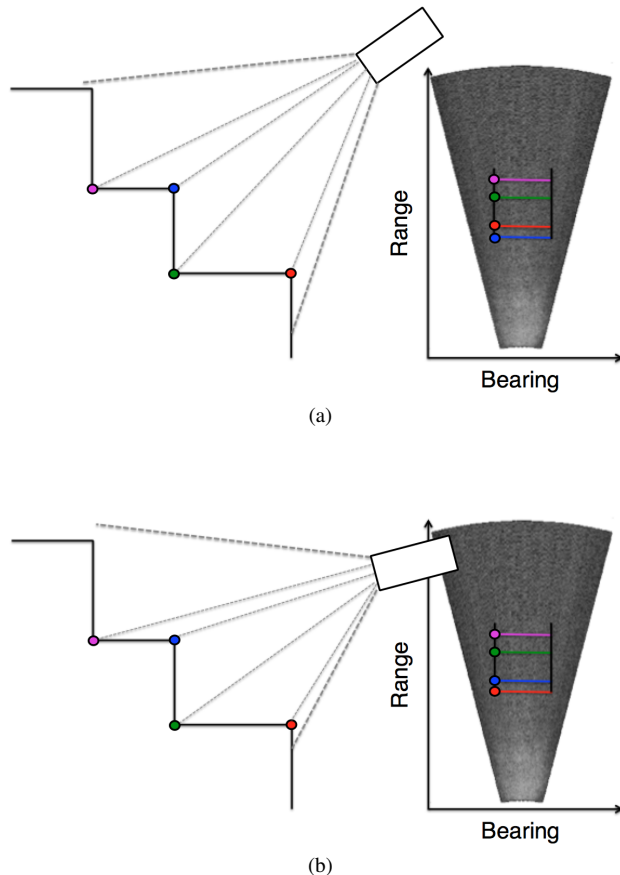


Fig. 1: Data association for imaging sonar presents several challenges. First, sonar images are very non-intuitive to interpret. The representation of structure in the image often does not correlate directly with the real 3D structure, as can be seen in (a) where the order of the colored points do not agree with our intuition based on visual imagery. Second, changing position and angle of the sonar could produce similar images but with different correspondences between feature points and real 3D points. The difference between (a) and (b) serves as an example. This complication makes even manual data association difficult.

SLAM methods for AUVs. First, unlike previous approaches, no planar assumptions about the environment are required. Second, ASFM is more accurate than pairwise approaches because it uses constraints from more than two images to eliminate drift and recover 3D landmark positions. Finally, ASFM does not have any of the limitations associated with cameras underwater such as shorter range or limited visibility in turbid water. Here, we build upon the bundle adjustment portion of the ASFM pipeline as described in [6] by adding incremental automatic data association.

Feature extraction and data association, or finding which measurements from different views correspond to the same object, make up the first part of the SLAM problem. Once

This work was partially supported by ONR grants N00014-14-1-0373, N00014-12-1-0093, and N00014-16-1-2365.

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

feature correspondences are known, the constraints can then be optimized to find the maximum likelihood set of robot poses and landmark positions. Data association is crucial because incorrect correspondences can drastically degrade the quality of the resulting map and trajectory.

To address the challenge of ensuring the data association is accurate and robust, we introduce a novel algorithm that to the best of our knowledge is the first to solve data association with an imaging sonar for this general case where the exact sonar pose is not known and the structure is not planar. The algorithm uses a tree of correspondences similar to that of Joint Compatibility Branch and Bound [8]. At each level of the tree, a feature measurement from the current image is matched with either one of the existing landmarks or with a “null” match, which means it becomes a new landmark. We use a gating threshold to prune the leaves of the tree, and each path down the tree from the top level to a leaf node is a possible data association hypothesis. Each hypothesis has an associated cost related to its posterior probability. We sort the hypotheses in order of increasing number of null matches, and then search through all paths from top level to leaf until we find one with a small enough cost. In other words, we return the set of correspondences with the fewest number of null matches and smallest posterior compatibility cost under a  $\chi^2$  threshold.

## II. RELATED WORK

Most of the related ideas on incremental data association for imaging sonar rely on a planar assumption of the scene. For instance, Leonard et al. [7] use a multiple hypothesis tracking (MHT) algorithm to perform data association and reconstruct the geometry of a static, rigid, 2D environment. Similar to our algorithm, MHT creates a tree of possible hypotheses matching measurements to features, and the tree is pruned based on the likelihood of each hypothesis. MHT in this work assumes that the landmark can be initialized with one measurement, which is not true for 3D scenes. Hover et al. [5] also require a planar assumption while using an AUV and forward looking sonar to perform ship hull inspection. For sonar image registration, they use a normal distribution transform (NDT) to provide a measure of the likelihood that a point will be observed by the sonar model.

Ribas et al. [11] use an individual compatibility test with a  $\chi^2$  threshold to determine which previous features could be correspondences. A nearest neighbor criterion is then applied to select the previously seen feature with the smallest Mahalanobis distance. Once again, the authors in this paper make a planar assumption by only imaging planar objects. In addition, the nearest neighbor criterion does not take into account the joint hypothesis of the entire set of features like our data association algorithm and therefore is more susceptible to accepting spurious features and producing incorrect data associations. Peillot et al. [10] present methods for 2D obstacle mapping and avoidance, the main application being surveys of the seabed. Unlike ASFM, which uses incremental smoothing, the authors use a Kalman filter to track objects in the forward-looking sonar images. The smoothing performed

by ASFM has an advantage over Kalman filters because it takes into account the entire history of the measurements and trajectory in the optimization, resulting in a more accurate map. In addition, even though there are more measurements and variables, the sparsity of smoothing makes it more efficient to compute than optimizing with the dense matrices of Kalman filters.

Fallon et al. [4] discuss a system that uses FLS sonar images to find and navigate to a previously mapped target. For data association, a scoring algorithm was used that takes into account positive information of features detected by the sonar and negative information of features that were expected to be seen but were not detected. Our work is similar to some of the ideas such as scoring SLAM graph hypotheses, but our data association applies to matching more generally with previous sonar images for 3D reconstruction instead of relying on a prior map. Fallon et al. also estimate whether a sonar feature is a moored or seafloor feature, but our ASFM covers a larger variety of landmarks at a continuous range of elevations. Aykin et al. [3] recover elevation from sonar images by utilizing the geometry of visual artifacts like shadows cast by stationary objects in the environment. They relax the planar assumption for pairwise matching of sonar frames but still assume a locally planar surface in order to include shadow information.

Note that all of the related work mentioned so far have only dealt with the 2D locations of sonar features, while our technique can deal with the more general 3D case. In the direction of 3D reconstruction, Assalih [1] introduces an acoustic stereo system that uses two sonars to recover 3D geometry. Given the measurement of a feature in one pose, the author estimates the measurement of the same feature in a different pose with a Modified Discrete Uniform Distribution, which samples points along the elevation arc and assigns probabilities of them being the true 3D sonar point. Unlike this method, ASFM only requires one sonar and can utilize information from more than two sonar images for 3D geometry recovery. More recent work by Aykin et al. [2] presents a space-carving method for recovering 3D geometry from multiple 2D forward-looking sonar images at known poses. Finding the closest edge of an object in multiple sonar images provides information about the occupancy of 3D voxels in the sonar field of view. This method achieves 3D reconstruction without the need for data association and feature extraction. However, ASFM constrains both the motion of the sonar as well as landmark positions, so unlike the space-carving method, the sonar poses do not need to be fully known a priori.

## III. INCREMENTAL DATA ASSOCIATION

### A. Sonar Geometry

Before discussing further how data association is performed, it is important to understand the information provided in a FLS sonar image. The imaging sonar sends out an acoustic ping and measures the intensity of acoustic waves reflected from objects, which is provided in the form of an intensity image. As seen in Fig. 2, the sonar only provides

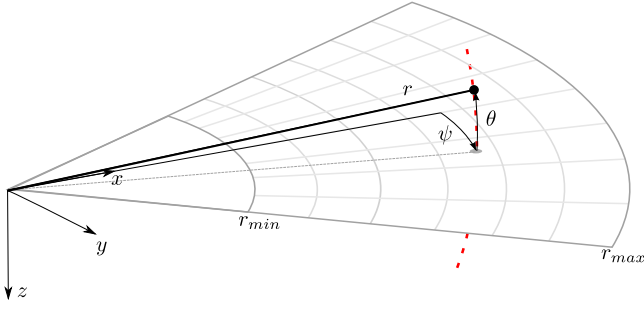


Fig. 2: Imaging sonar geometry. Any 3D point along the dashed red elevation arc will appear as the same image point in the  $x$ - $y$  plane. Range  $r$  and bearing angle  $\psi$  are measured, but the elevation angle  $\theta$  is lost in the projection process.

partial information about a feature (bearing  $\psi$  and range  $r$ ) and does not provide its elevation angle  $\theta$ . In a 1-D array of receivers, the difference between the time it takes for one receiver to gather a signal and another receiver to gather the same signal denotes the bearing of the feature. The range is determined by the time of flight of the sound signal. The elevation of the point is lost, as all points along an elevation arc within the field of view of the sonar will collapse to the same pixel in the sonar image. Since one dimension of the feature is missing, one sonar image is not sufficient to recover 3D geometry.

### B. ASFM Overview

Since this paper focuses on the data association aspects of ASFM, only a brief overview is given of the ASFM problem. See [6] for more details on the ASFM factor graph and optimization. The goal of acoustic structure from motion is to find the most likely set of vehicle poses and 3D landmark positions given odometry and sonar measurements. We first define a generative sensor model using Gaussian noise:

$$\begin{aligned} g(x_{i-1}, x_i) + \mathcal{N}(0, \Lambda_i) \\ h(x_i, l_j) + \mathcal{N}(0, \Xi_k) \end{aligned} \quad (1)$$

where  $g(x_{i-1}, x_i)$  is the odometry measurement prediction between pose  $x_{i-1}$  and pose  $x_i$  and  $h(x_i, l_j)$  is the sonar measurement prediction of landmark  $l_j$  observed from pose  $x_i$ .  $\Lambda_i$  is the covariance of the odometry measurement between pose  $x_{i-1}$  and pose  $x_i$ , while  $\Xi_k$  is the covariance of the  $k^{th}$  sonar measurement. The sonar measurement prediction  $h(x_i, l_j)$  first transforms the 3D landmark location  $l_j = (x, y, z)$  in the global frame to the local pose  $x_i$  frame, obtaining the local coordinates  $(x_s, y_s, z_s)$ . Bearing  $\psi$  and range  $r$  are then obtained by

$$r = \sqrt{x_s^2 + y_s^2 + z_s^2} \quad (2)$$

$$\psi = \text{atan2}(y_s, x_s). \quad (3)$$

Using this sensor model, ASFM becomes a standard non-linear least squares problem:

$$\begin{aligned} \Theta^* = \underset{x}{\operatorname{argmin}} [ & \|x_0\|_{\Lambda}^2 + \sum_{k=1}^m \|h(x_i, l_j) - m_k\|_{\Xi_k}^2 \\ & + \sum_{i=1}^n \|g(x_{i-1}, x_i) - u_i\|_{\Lambda_i}^2 ] \end{aligned} \quad (4)$$

where  $\|x\|_{\Sigma}^2 = x^T \Sigma^{-1} x$  is the squared Mahalanobis distance. The first pose  $x_0$  is fixed at the origin using a prior,  $m_k$  is the  $k^{th}$  sonar measurement involving pose  $x_i$  and landmark  $l_j$ , and  $u_i$  is the  $i^{th}$  odometry measurement between poses  $x_{i-1}$  and  $x_i$ .

We initialize 3D landmark locations using the following backprojection function ( $r$  is the range and  $\psi$  is the bearing):

$$\begin{bmatrix} x_s \\ y_s \\ z_s \end{bmatrix} = r \begin{bmatrix} \cos \psi \cos \theta \\ \sin \psi \cos \theta \\ \sin \theta \end{bmatrix} \quad (5)$$

where the unknown elevation,  $\theta$ , is set to 0. The non-linear least squares problem including the sonar poses and 3D landmark positions is then solved using iterative linearization with Powell's dogleg or Levenberg-Marquardt.

### C. Data Association Challenges

Unlike camera images, sonar images are much less intuitive to understand and interpret. An example is given in Fig. 1. Assume the AUV is imaging a stair-like structure underwater and we have manually picked out some point features that intuition would lead us to believe are stable, like the corners along one edge of the stairs. Since the vertical axis of the sonar image is range from the sonar and the blue feature appears to be closest to the sonar, the blue point appears as the bottom-most feature in the sonar image. The next closest point to the sonar looks to be the red feature, then the green, then the purple. Note that just looking at the final ordering of the feature points in the sonar image does not give a helpful indication of the true 3D structure. From only the sonar image, it would be almost impossible to tell that in true 3D, the blue feature point is in fact between the green and the purple feature points.

To confuse data association further, moving and rotating the sonar changes the ordering of the feature points because the distance between the features and the sonar changes. Therefore, in (b) of Fig. 1, the sonar moves and the resulting sonar image contains a different ordering of feature points. In this case, the sonar moves closer to the red feature point and further from the blue feature point. Consequently, the blue and red features switch places in the second sonar image. Without knowing the exact motion of the AUV, even manually assigning feature correspondences becomes difficult. It would be very challenging to correctly associate the blue feature point in the first image to the blue feature point in the second image. Given the sonar projection geometry, not only manual data association, but also manually picking features that are stable and can be seen by multiple frames in a variety of viewing poses is not an easy task.

### D. Incremental Data Association Algorithm

Our algorithm proceeds incrementally with every new sonar frame and odometry measurement. To initialize the algorithm, the first pose in the data sequence is fixed and set as a prior in the ASFM factor graph. Additionally, all landmark measurements from the first pose are regarded as new landmarks and stored in a landmark history. Next,

incrementally, a new odometry measurement will arrive along with a set of feature measurements from this new pose. If feature measurements are too close to each other, we will discard them to avoid ambiguity. In our experiments, if two measurements were within  $1^\circ$  bearing and 0.2 m range of each other, both were discarded.

For each new feature measurement, we look at the reprojection error of each landmark in the stored landmark history to prune possible matches. If a landmark has only been seen once so far, it is backprojected through the pose that saw the landmark before into 1-degree interval 3D points along the elevation arc of possible 3D points (Fig. 2). These 3D points are then reprojected into the current pose and the landmark is accepted as a possible match if any of its associated reprojected 3D points lies within a gating threshold of the current feature measurement. If a landmark has been seen twice or more so far, its current 3D location estimate from the optimization is used as the 3D point that is reprojected into the current pose. The gating threshold is empirically set to a 0.2 m range difference and  $4^\circ$  bearing (0.07 radian) difference between the reprojected landmark and the current feature measurement in our experiments. To avoid keeping unnecessary landmarks, a new landmark is removed from the stored landmark history if it isn't seen again, and therefore cannot be fully constrained, in the next two sonar frames being used for ASFM.

Once all possible matches have been determined for each new feature measurement using the gating threshold, a search over possible match combinations after pruning is utilized to find the correct data association hypothesis. More specifically, we build a tree where each level represents a feature observation and its potential matches (including a null match) as determined by the gating threshold. Each path from a node in the top level to a leaf node at the bottom level represents a potential data association hypothesis. The data association hypothesis is accepted if the sum of squared residuals for the posterior position of the landmarks and the robot pose given this hypothesis falls under a  $\chi^2_{d,\alpha}$  threshold:

$$\|x_0\|_{\Lambda}^2 + \sum_{k=1}^m \|h(x_{i_k}, l_{j_k}) - m_k\|_{\Xi_k}^2 + \sum_{i=1}^n \|g(x_{i-1}, x_i) - u_i\|_{\Lambda_i}^2 < \chi^2_{d,\alpha} \quad (6)$$

Feature measurements and landmarks are not added to the optimization if they are not fully constrained, i.e. they have not been seen by at least two different poses. The  $\chi^2_{d,\alpha}$  threshold is determined using  $d$ , the degrees of freedom of the factor graph (number of measurements minus number of variables) and  $\alpha$ , the confidence parameter (set to 0.99 in our experiments). Null matches, or declaring feature measurements to be new landmarks, are penalized such that the algorithm picks the hypothesis that fits the  $\chi^2_{d,\alpha}$  criterion with the fewest null matches. If there are multiple hypotheses that fit the  $\chi^2_{d,\alpha}$  criterion with the same number of null matches, the algorithm picks the hypothesis with the smallest optimization residual. In the worst case, the algorithm has a

computational complexity of  $O(ml^m \log(l))$  where  $m$  is the total number of sonar features seen in the current pose and  $l$  is the total number of landmarks in the landmark history.

A downside of our algorithm is that the posterior of the entire set of landmarks and vehicle poses has to be calculated for each hypothesis. However, before performing the restricted search over viable candidates after pruning, the data association hypotheses are sorted in order of increasing number of null matches. Therefore, if small sonar motion is assumed, which is not necessary for the algorithm to work, but applicable to many situations, then many features will overlap. The correct data association hypothesis will then contain few null matches and will be found quickly.

## IV. EXPERIMENTAL RESULTS

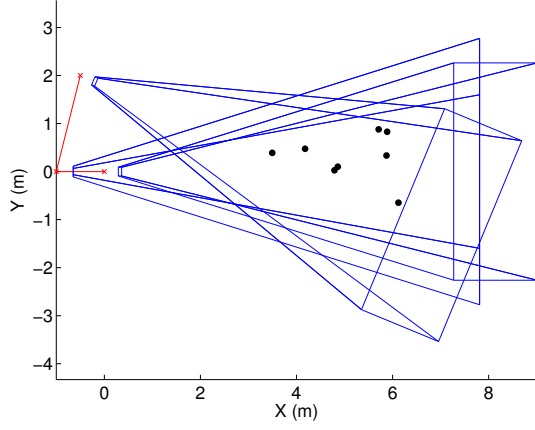
### A. Simulation

We present statistical results for general vehicle motion using simulated data. The simulation data was generated by randomly creating three sonar poses and 3D points until at least eight 3D points were visible in all three sonar frames. The number of features used was empirically determined as a realistic number of features that could be seen by three sonar poses of varying viewing angle. Gaussian noise was added to the bearing ( $\sigma = 0.2^\circ$  for small noise,  $\sigma = 0.5^\circ$  for larger noise) and range ( $\sigma = 0.005$  m for small noise,  $\sigma = 0.01$  m for larger noise) components of the ground truth sonar measurements. Gaussian noise was also added to both rotational ( $\sigma = 1^\circ$ ) and translational components ( $\sigma = 0.01$  m) of the odometry between consecutive poses. The simulated sonar and environment specifications are listed in Table I. Ten different sonar and point environments were randomly generated for the simulation experiments. Examples of environments are shown in Figs. 3 and 4.

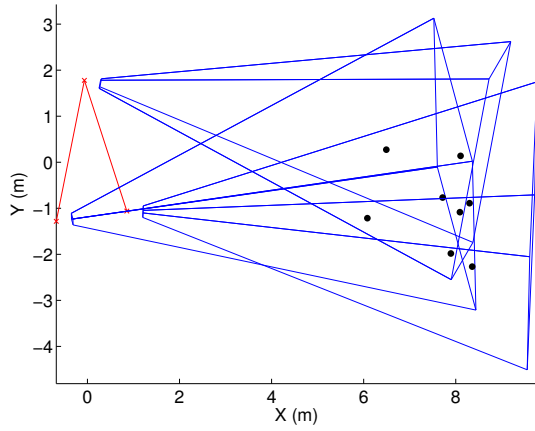
Some important practical considerations relate to ambiguity and local minima. To avoid ambiguity, we discard both measurements if a feature is within  $1^\circ$  bearing or 0.2 m range of another feature measurement. Without this additional threshold, the incorrect data association confusing these features often has a similarly small, if not better, posterior than the correct hypothesis. Additionally, we found that it is important to set up the factor graph in a batch manner when testing a hypothesis. If the graph is built incrementally, meaning when new measurements arrive, they are used to update an existing solution, the optimization is more likely to get trapped in a local minimum. The new

TABLE I: Simulated data experimental design

	Value
Number of Monte Carlo samples	100
Orientation: stddev (deg)	1
Translation: stddev (m)	0.01
Bearing: stddev (deg)	0.2 (small), 0.5 (large)
Range: stddev (m)	0.005 (small), 0.01 (large)
Minimum range of sonar (m)	0.375
Maximum range of sonar (m)	9.375
Bearing FOV of sonar (deg)	28.8
Elevation FOV of sonar (deg)	28
Number of bearing bins	96
Number of range bins	512



(a) Environment 2



(b) Environment 4

Fig. 3: Top views of examples of a random environment generated for the Monte Carlo simulations. The red line shows the trajectory of the sonar, the blue frustums demonstrate the frustum fields of view of the sonars, and the black dots are the eight 3D points that can be seen by all three sonar poses.

measurements can provide essential information to resolve previously ambiguous situations, and starting from a neutral initialization point for the elevation angles makes it less likely to get stuck in a local minimum.

We use Monte Carlo sampling to analyze the accuracy and robustness of our incremental data association algorithm. Four experiments were performed: small and large noise, including spurious feature measurements and without spurious features. Each sonar trajectory was simulated 100 times with noise randomly sampled each time from the same Gaussian distribution ( $\mu = 0$  and  $\sigma$  as described above). For each pose in each trial, five of the eight 3D points that could have been seen were randomly chosen to be measured. Consequently, not all of the poses saw all of the same points and the correct data association hypothesis often contained several null matches. For the spurious feature experiment, a random number of spurious features (0 to 2) were added for each pose. The spurious features were generated by randomly creating measurements in the sonar field of view.

The average accuracy of our data association algorithm

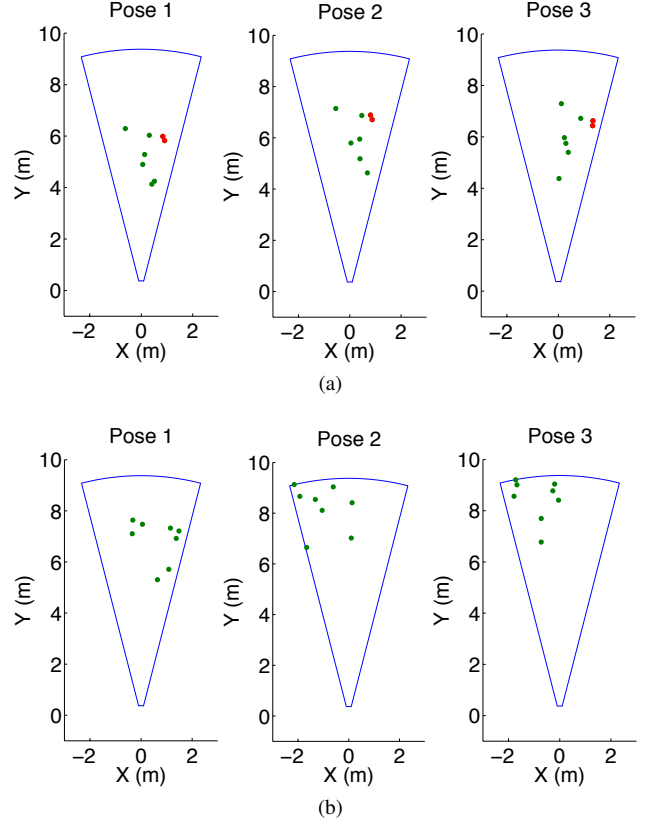


Fig. 4: Feature measurements for environments (a) 2 and (b) 4. In environment 2, the two red points were deemed too close, so both points were discarded to avoid ambiguity.

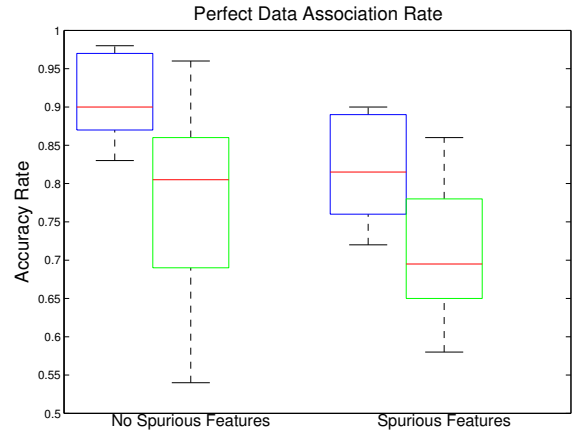


Fig. 5: Rate at which our data association algorithm found the exact correct hypothesis with no mistakes. The mean accuracy rate with small noise (blue) without spurious features was 0.91 with a standard deviation of 0.05, while with larger noise (green) it was 0.79 with a standard deviation of 0.13. The mean accuracy rate including spurious features for small noise was 0.82 with a standard deviation of 0.07, while for larger noise it was 0.72 with a standard deviation of 0.09.

over 100 runs for each of 10 environments without the inclusion of spurious features for small noise ( $\sigma = 0.2^\circ$ ,  $\sigma = 0.005$  m) was 90.9%, meaning that the algorithm found the exactly correct data association 90.9% of the time, while for larger noise ( $\sigma = 0.5^\circ$ ,  $\sigma = 0.01$  m) the average accuracy was 78.9%. With spurious measurements, the accuracy drops to 81.6% for small noise and 72.4% for larger noise (Fig. 5). Many of the errors were caused by the algorithm being a little more conservative and finding a data association hypothesis that had more null matches than the correct one. Tuning the  $\chi^2_{d,\alpha}$  threshold confidence level  $\alpha$  can change how conservative the algorithm is. A lower  $\alpha$  increases the percentage of the distribution that will be thrown out as outliers, thereby increasing the likelihood that a correct hypothesis will be rejected.

A box plot of the squared and weighted landmark factor residuals with and without spurious features for small noise is shown in Fig. 6. The landmark factor residual is the reprojection error of the final 3D point recovered and the original feature measurement. Note that the squared landmark factor residual is weighted by the inverse covariance matrix, so it becomes a unitless value. The average landmark factor residual over all trials without spurious features was 0.55 with a standard deviation of 0.29. Including spurious features, the average reprojection error was 0.47 with a standard deviation of 0.29. As expected, the addition of spurious features does not affect the landmark residuals very much because the spurious features should rarely ever be added to the factor graph as they do not correspond to real landmarks.

One of the environments that consistently has lower accuracy rates is environment 4. As shown in Fig. 3, the sonar motion between poses is quite large. Compared to a more well-behaved environment such as environment 2, the largest motion between poses in environment 4 is about 0.5 m larger in the  $x$ -direction, 0.7 m larger in the  $y$ -direction, and has about  $40^\circ$  more change in roll. These larger changes in sonar motion between frames could contribute to more ambiguity amongst features even if they do not appear very close to each other because the geometry and correlation of the features can change significantly between more radically different points of view. In a real application, it is usually safe to assume that the sonar motion between frames is small because for mapping purposes AUVs typically move very slowly. For instance, the Bluefin Hovering Autonomous Underwater Vehicle (HAUV) we use in our real data experiments usually travels at speeds of about 0.3 m/s.

Runtime of our algorithm over the different randomly generated environments with small noise is shown in Fig. 7. The order of the inputs into the search reduces the runtime significantly given that we favor hypotheses with fewer null matches. Without spurious features, the average runtime using a C++ implementation on a 2.5 GHz Intel i7 processor was 14ms with a standard deviation of 23ms. Including spurious features, the average runtime was 16ms with a standard deviation of 25ms. Environment 9 stands out in terms of longer runtime because the feature points were

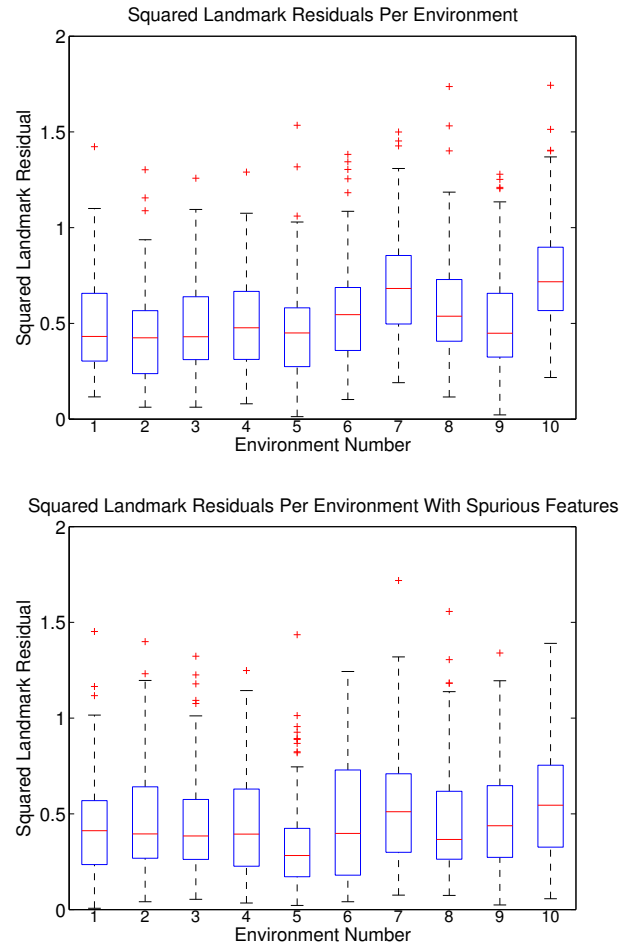


Fig. 6: Box plots of individual squared landmark residuals (weighted reprojection error) per environment. The average squared landmark residual (unitless) over all environments for small noise without spurious features was 0.55 with a standard deviation of 0.29. Including spurious features, the average landmark residual was 0.47 with a standard deviation of 0.29.

relatively close together. Therefore, each feature matched up to many of the landmarks in the landmark history and more hypotheses needed to be tested to find the correct one.

### B. Imaging Sonar Sequence

We demonstrate 3D structure recovery with incremental data association from several imaging sonar frames recorded with a Bluefin Hovering Autonomous Underwater Vehicle (HAUV) (Fig. 8) in Boston, Massachusetts. Five sonar frames were selected from the dataset to perform ASFM and point features were manually selected from all five sonar frames. We also randomly generated a random number of spurious features (0 to 2) in each sonar image to test the algorithm's robustness on real data (Fig. 9). Although features were extracted manually, point correspondences were found automatically using our data association algorithm.

In our experiments, we use a Sound Metrics DIDSON 300m forward-looking sonar [12]. It has a  $\psi_{max} = 28.8^\circ$  bearing field of view (FOV) and a  $28^\circ$  vertical FOV (using a spreader lens). The DIDSON sonar discretizes returns into  $w = 96$  bearing bins and  $h = 512$  range bins. The DIDSON mode used for this dataset provides a minimum range of

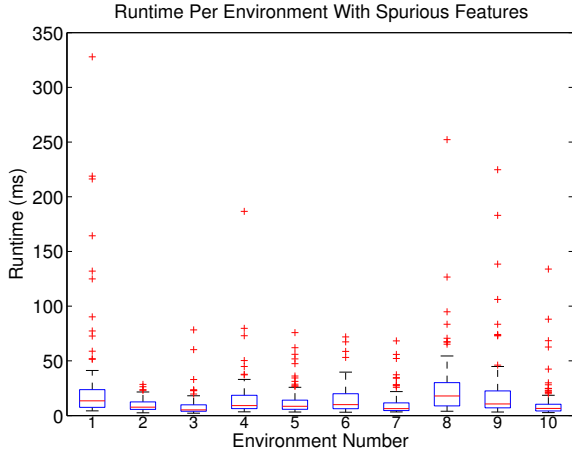
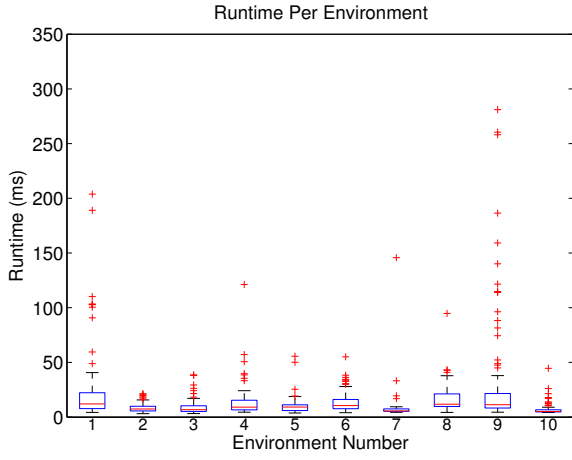


Fig. 7: Box plots of runtime in milliseconds per environment for the data association algorithm using a C++ implementation on a 2.5GHz Intel i7 processor. The average runtime over all environments for small noise without spurious features was  $14ms$  with a standard deviation of  $23ms$ . With spurious features, the average runtime was  $16ms$  with a standard deviation of  $25ms$ .

$r_{min} = 0.75$  m and a maximum range of  $r_{max} = 5.25$  m. Let  $(u, v)$  be the image coordinates of a feature in the Cartesian sonar image, and  $\gamma$  be a constant describing the number of pixels per meter in the Cartesian image. Since we extract features from Cartesian space, the bearing  $\psi$  and range  $r$  are obtained using:

$$\gamma = \frac{w}{2r_{max}\sin(\frac{\psi_{max}}{2})} \quad (7)$$

$$x_s = \frac{u - \frac{w}{2}}{\gamma} \quad (8)$$

$$y_s = r_{max} - \frac{v}{\gamma}. \quad (9)$$

$$r = \sqrt{x_s^2 + y_s^2} \quad (10)$$

$$\psi = \text{atan2}(x_s, y_s). \quad (11)$$

$\psi$  additionally needs to be corrected with a lens distortion function provided by the manufacturer. Odometry readings from the vehicle were used in the optimization to further constrain the problem. The odometry was obtained from a ring-laser gyro and a Doppler Velocity Log (DVL). We chose odometry uncertainties of  $\sigma = 1^\circ$  for rotation and  $\sigma = 0.1$

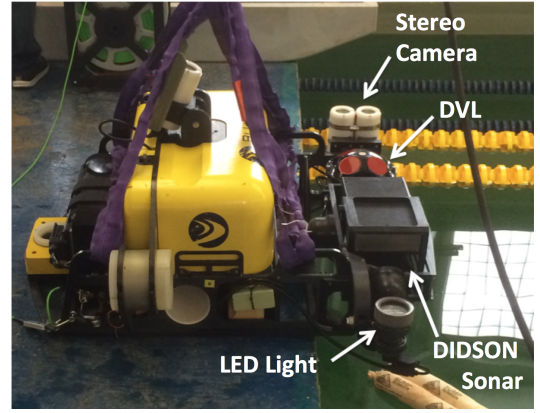


Fig. 8: Bluefin Hovering Autonomous Underwater Vehicle (HAUV) used in our real data experiments. The DIDSON sonar and Doppler Velocity Log (DVL) are pictured attached to the front of the vehicle.

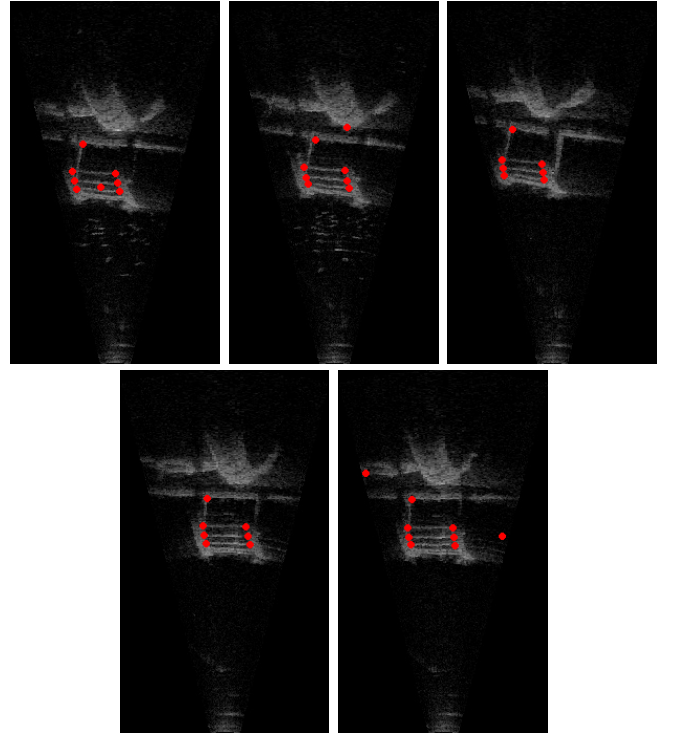


Fig. 9: Manually marked features (red circles) for the five raw sonar frames that were used to reconstruct the ladder geometry with the addition of 0 to 2 randomly generated spurious features. The sonar images show the entire Cartesian FOV of the sonar (height of image spans 4.5 m).

m for translation. For bearing and range measurements from the DIDSON sonar we use  $\sigma = 0.2^\circ$  and  $\sigma = 0.005$  m respectively.

The 3D geometry of the ladder in the imaging sonar sequence was recovered as shown in Fig. 10. The algorithm correctly ignored the spurious features and found the true hypothesis in  $232ms$ . Without spurious features, we were also able to generate the correct data association in  $230ms$ . Before optimization, the ladder is initialized as a flat object with no elevation. The structure in the  $x-y$  plane looks convincing, but from the  $x-z$  view, it is clear that the initialization does not capture the reality that the ladder's rungs are

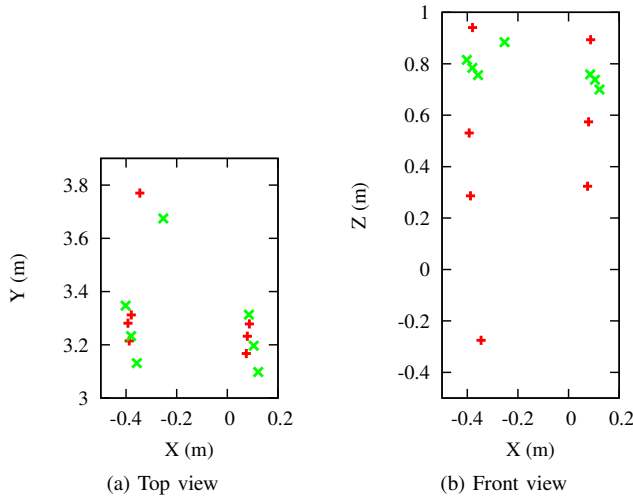


Fig. 10: (a) Top and (b) front views of 3D ladder structure before (green 'x') and after (red '+') optimization from five imaging sonar frames. While initially the ladder's rungs are collapsed in elevation, ASFM recovers the actual structure, albeit with some remaining noise.

at different  $z$  elevations. Since the data association algorithm was successful in finding the correct point correspondences, we get the same results as described in our previous work [6]. The difference in this result is that we were able to discard spurious features as outliers and everything besides the feature extraction was done automatically.

## V. CONCLUSION

We have presented a novel incremental data association algorithm for finding point correspondences between multiple 2D sonar images of nonplanar structures. To our knowledge, no other data association method currently exists for 3D reconstruction of general scenes using an imaging sonar and uncertain sonar poses. Simulations of randomly generated sonar trajectories show the ability of our algorithm to find the correct data association hypothesis with a high success rate. The inclusion of spurious measurements in our simulation experiments further demonstrates the robustness of our data association. An experiment with real sonar data containing spurious features and manually extracted feature points shows the successful incorporation of the algorithm into the ASFM pipeline for 3D geometry recovery.

A possible improvement to our incremental data association algorithm is the use of the Incremental Posterior Joint Compatibility Test (IPJC) [9], which uses the same ideas as our current algorithm by searching a tree of hypotheses and computing a posterior compatibility cost. However, IPJC approximates the  $\chi^2_{d,\alpha}$  error with an Extended Kalman Filter (EKF) update step instead of using a full optimization. If the correct data association most often includes very few

null matches, our pruned search would already be very fast. Nevertheless, more generally if there are many spurious features or not many overlapping features with recent sonar poses, IPJC could reduce the computational time of our algorithm significantly.

Of course, the automatic ASFM pipeline would not be complete and practical for real-time applications without an automatic feature extractor. What kind of features are most stable and useful in imaging sonar images remains an open problem. Many computer vision features have been developed for cameras, but they might not be the best fit for the unique projective geometry of the sonar. Further research is needed to determine suitable features for sonar images.

## ACKNOWLEDGMENTS

The authors would like to thank Dr. Jason Stack for his support, Pedro Vaz Teixeira for recording the ladder sequence, and the anonymous reviewers for their valuable comments.

## REFERENCES

- [1] H. Assalih, "3D reconstruction and motion estimation using forward looking sonar," Ph.D. dissertation, Heriot-Watt University, 2013.
- [2] M. Aykin and S. Negahdaripour, "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.
- [3] —, "On feature matching and image registration for two-dimensional forward-scan sonar imaging," *J. of Field Robotics*, vol. 30, no. 4, pp. 602–623, Jul. 2013.
- [4] M. F. Fallon, J. Folkesson, H. McClelland, and J. J. Leonard, "Relocating underwater features autonomously using sonar-based SLAM," *IEEE J. Ocean Engineering*, vol. 38, no. 3, pp. 500–513, Jul. 2013.
- [5] F. Hover, R. Eustice, A. Kim, B. Englot, H. Johannsson, M. Kaess, and J. Leonard, "Advanced perception, navigation and planning for autonomous in-water ship hull inspection," *Intl. J. of Robotics Research*, vol. 31, no. 12, pp. 1445–1464, Oct. 2012.
- [6] 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.
- [7] J. J. Leonard, B. A. Moran, I. J. Cox, and M. L. Miller, "Underwater sonar data fusion using an efficient multiple hypothesis algorithm," in *Proc. IEEE Int. Conf. Robotics and Automation*, May 1995, pp. 2995–3002.
- [8] J. Neira and J. D. Tardos, "Data association in stochastic mapping using the joint compatibility test," *IEEE Trans. Robotics and Automation*, vol. 17, no. 6, pp. 890–897, Dec. 2001.
- [9] E. Olson and Y. Li, "IPJC: The incremental posterior joint compatibility test for fast feature cloud matching," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, Oct. 2012, pp. 3467–3474.
- [10] Y. Petillot, I. T. Ruiz, and D. M. Lane, "Underwater vehicle obstacle avoidance and path planning using a multi-beam forward looking sonar," *Journal of Oceanic Engineering*, vol. 26, pp. 240–251, Apr. 2001.
- [11] D. Ribas, P. Ridao, J. Neira, and J. Tardós, "SLAM using an imaging sonar for partially structured underwater environments," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, Oct. 2006, pp. 5040–5045.
- [12] Sound Metrics Corporation, "SoundMetrics Didson 300 specifications," <http://www.soundmetrics.com/products/DIDSON-Sonars/DIDSON-300-m/>.