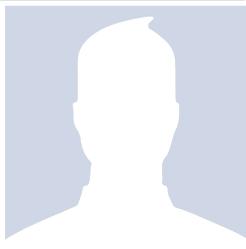


# Right on target



Peter King



Benjamin Anstey



Andrew Vardy

*King, Anstey, and Vardy apply image-processing techniques to AUV-collected sonar data for the purposes of enhanced localization.*

## **Who should read this paper?**

Those with an interest in AUVs, robotics, or sonar data processing.

## **Why is it important?**

AUVs are becoming ever more present in the ocean industry. This paper presents an example of the type of techniques that are being developed to help them perform better and become more “autonomous.” Specifically, it discusses a system to provide augmented localization to an AUV equipped with a side scan sonar. The innovation is in the methodology used to match images collected at different times, but of the same area. Through some simple assumptions, based on the characteristics of the AUV, the matching becomes quite robust. Augmented localization helps improve navigation performance of AUVs and of the data quality they return as well as leading to more robust operations.

## **About the authors**

Peter King is the AUV Facility Coordinator for the Australian Maritime College, University of Tasmania, currently working under the Antarctic Gateway Partnership Project, a Special Research Initiative of the Australian Research Council. His work focuses on the operations and development of AUVs, specifically in dynamic areas, such as beneath ice. His areas of expertise include automated sonar image generation and processing, and how they can be used to drive autonomous navigation. Mr. King completed his B.Eng. (Computer) at Memorial University of Newfoundland and currently is a part-time PhD candidate.

Benjamin Anstey is a software developer for Verafin in St. John’s, NL, Canada. Previously he was a research engineer with Memorial University’s Marine Environmental Lab for Intelligent Vehicles. He has a B.Eng. (Computer) from Memorial University.

Dr. Andrew Vardy completed a B.Eng. in electrical engineering (computers and communications option) from Memorial University in 1999. After graduating, he became interested in bio-inspired computing and pursued this interest by completing a master’s degree from the University of Sussex. He continued his studies by completing a PhD in computer science from Carleton University. Dr. Vardy returned to Memorial University in 2005, jointly appointed to the Department of Computer Science and the Faculty of Engineering and Applied Science. His main research interest is biologically inspired robotics. He is currently working on swarm robotics and autonomous navigation for underwater vehicles. He teaches courses in autonomous robotics, computer programming, and control systems.

# SONAR IMAGE REGISTRATION FOR LOCALIZATION OF AN UNDERWATER VEHICLE

**Peter King<sup>1</sup>, Benjamin Anstey<sup>2</sup>, and Andrew Vardy<sup>3</sup>**

*<sup>1</sup>Australian Maritime College, University of Tasmania, Launceston, Australia*

*<sup>2</sup>Verafin Ltd., St. John's, NL, Canada*

*<sup>3</sup>Department of Computer Science, Department of Electrical and Computer Engineering, Memorial University of Newfoundland, St. John's, NL, Canada*

## ABSTRACT

This paper presents a system to provide augmented localization to an AUV equipped with a side scan sonar. Upon revisiting an area, from which side scan data had previously been collected, the system generates an estimate to bound the error in the AUV's estimate. Localization is accomplished through the comparison of sonar images.

Image comparison is based on the extraction of features which characterize local gradient distributions, such as Lowe's SIFT feature extractor. To resolve potential ambiguities and noise in the image comparison measurement, the localization system incorporates a Bayesian inference algorithm that considers both image based measurement and relative motion to refine the position estimate over time. We describe the particular methods, constraints and augmentations used to apply established image matching and alignment techniques to side scan sonar imagery. By applying consistent geographical corrections to the raw sonar data; using a flat-bottom assumption; and by adding the constraint that images are formed with north aligned up; the traditional problem of full pose estimation is reduced to the two-dimensional case of determining only the  $x,y$  translation independent of vehicle altitude. Due to the assumption of constant scale and orientation between images, sensitivity of image feature matching is shown to be controllable by filtering feature matches based on comparing their scale and orientation. This effect was quantified using binary classification analysis. The system's performance was measured by performing tests on a large side scan survey which represents the familiar terrain that a returning AUV could use for localization.

## INTRODUCTION

Autonomous underwater vehicles (AUVs) employ a host of different techniques for localization, but there remain many operational scenarios when these techniques prove insufficient or require augmentation. One example is long endurance missions when either no support vessel is in close proximity or access to the surface for satellite positioning is infeasible or undesirable.

One of the many risks associated with under-ice missions is that the AUV may be unable to return precisely to its launch and recovery point due to the unavoidable growth of navigational drift in its position estimate. In deep water missions, returning to the surface to obtain a satellite-aided position fix is possible, but should ideally be avoided to devote more time to executing the desired survey. Deep water surveys have the additional challenge of increased uncertainty during descent and ascent when ground lock from Doppler velocity log (DVL) sensors is unavailable. In these cases, we propose that an AUV equipped with side scan or other imaging sonar could perform targeted surveys of the seabed around locations where it is anticipated that augmented localization would be beneficial. Upon revisiting the surveyed region with increased navigational drift error, the proposed system would be employed to generate a new corrected position estimate. In brief, the proposed system compares the current sonar image of the seabed with previously surveyed images yielding a belief estimate that is refined over time to reduce positional ambiguity. Our particular approach is based on the extraction of features which characterize local gradient distributions, such as Lowe's SIFT feature extractor [Lowe, 2004].

This paper makes contributions in the area of image-based localization which incorporates Bayesian inference to help resolve potential ambiguities in the unfiltered sensor data. We present results on images collected by Memorial University's Explorer AUV, built by International Submarine Engineering of Port Coquitlam, British Columbia, Canada.

The particular challenges of underwater image-based localization (IBL) come from the unstructured nature of seabed terrain and the issues inherent in the imaging process. The imaging issues depend on the technology employed, typically optical or acoustic. Optical imaging, the default choice in mobile robotics, is constrained in its applicability by water turbidity (e.g., "marine snow") and illumination. External lighting is usually required which reduces the effective operational altitude of the sensor above the seabed due to the high level of attenuation of light in water. Acoustic imaging technologies such as side scan sonar can ensonify very large regions of the seabed, on the order of hundreds of metres. This increased coverage results in more expansive and efficient surveys over optical imaging. In acoustic imaging, pulses of acoustic energy known as pings are projected in a flattened fan shape to ensonify the target (the seabed or other underwater feature). A transducer samples the return intensities in time and this data is then projected into an image. Acoustic ranging technologies such as single- and multibeam differ in the number and shape of emitted pings, and in how they are interpreted. However, multibeam sonar can be used to produce images in addition to range and intensity data. Whatever the technology,

acoustic techniques for seabed imaging share some common issues. To produce an image which is appropriately scaled for physical interpretation, the seabed is often assumed to be flat (although the image can be overlaid on bathymetry data if available). Some estimate or assumed value for the speed of sound through the local water column must be employed. Also, height deviations in the seabed will often result in acoustic shadows which are essentially gaps where no acoustic energy was returned because of occlusion.

We focus here on side scan sonar, a mature technology that has been integrated into a wide variety of AUVs. The sensor consists of two transducer heads, mounted on either side of the AUV or towed body. Each transducer emits an acoustic ping projected towards the seabed in a fan shape oriented orthogonally to the sensor's direction of travel (the sensor's direction of travel is referred to as "along track," while the orthogonal direction is "across track"). Typically, the return intensities for each ping are plotted in a "waterfall" image which depicts the return intensities for each transducer, plotted against horizontal time

axes. See Figure 1 for an example. In the centre of the waterfall image is the nadir, which represents an effective gap in coverage due to the angle at which the sonar's transducers are mounted (to the sides). The nadir is often left as a void in the image or is otherwise disregarded.

Care must be taken in interpreting a waterfall image such as that shown in Figure 1. Pixels in this image cannot be interpreted as having a real and consistent physical scale. Turns of the vehicle or alterations in its velocity while sampling will yield increased distortion which will impair any type of manual or automated interpretation. There are a range of corrections and procedures required to generate images whose pixels can be interpreted as having a real physical scale and which are free of distortions induced by the motion of the vehicle. Some researchers have attempted to apply automated techniques to waterfall images [Stadler et al., 2008] but we believe these techniques would be more effective on corrected images. One of the contributions of this work is the image generation procedure which carries out the necessary corrections and

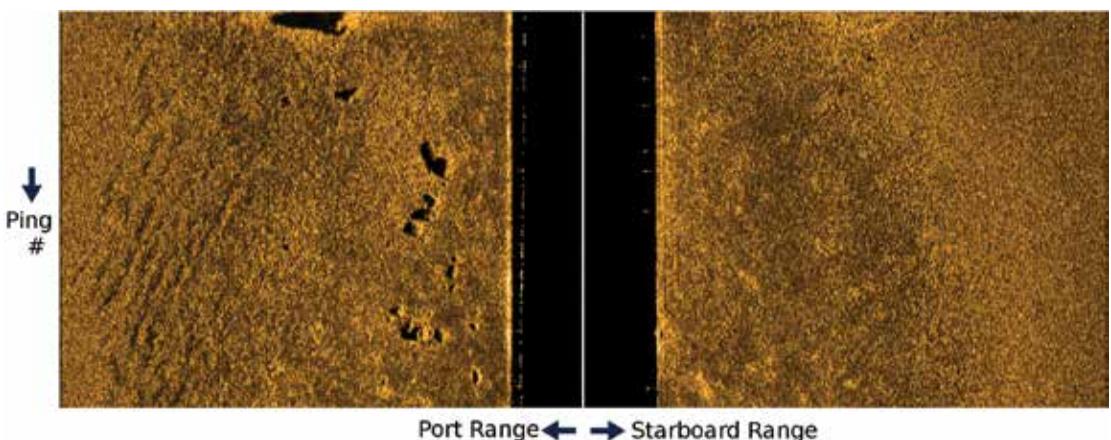


Figure 1: Example of a waterfall side scan sonar image.

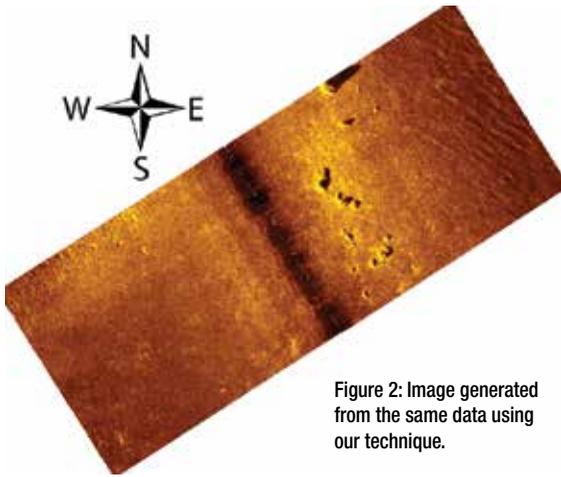


Figure 2: Image generated from the same data using our technique.

adjustments to produce high-quality images suitable for matching and localization, such as the one shown in Figure 2. Our image generation methods have been described in previous publications [King, 2014; King, 2012; Vandrish, 2012]. It is important to note that our image generation procedures run on the AUV itself in real-time, whereas conversion from raw side scan data to corrected images is traditionally done off-line by a trained expert using specialized software. Thus, our image generation procedure represents a substantial contribution.

Our localization system assumes that a surveyed area is subsequently revisited, potentially along a trajectory that differs from the initial survey. In order to facilitate subsequent matching, registration, and localization, we extract keypoints from the image. A keypoint can be defined as any distinctive image sub-region. We utilize techniques that inherit from Lowe's Scale-Invariant Feature Transform (SIFT) concept [Lowe, 2004]. These techniques, such as SURF [Bay et al., 2008] and FREAK [Alahi et al., 2012] characterize an image sub-region based on the local gradients within the sub-region, extracted at a scale particular to the

keypoint. Each keypoint is associated with a descriptor vector and the distance between the  $n$ -dimensional descriptor vectors relates to the similarity between image sub-regions.

In previous work, we explored a variety of keypoint extractors, descriptor vectors, and matching schemes [King et al., 2013]. A key contribution of our work in this paper is the exploration of the parameter space for keypoint extraction and matching on side scan sonar images.

Our work falls under the classification of image-based localization (IBL). A well-known issue in IBL is perceptual aliasing – different locations yield a similar appearance. This problem is quite pronounced when dealing with seabed images, which often lack significant texture and distinguishing patterns. If the entire surveyed region is lacking in texture and variation, then our scheme will not be successful. However, we assume that a completely featureless terrain is unlikely. As long as some of the surveyed area contains distinguishing patterns then localization should be possible. We use an implementation of the discrete Bayes filter [Thrun et al., 2005] to combine predictions from the previous belief model with the probability of obtaining the currently-perceived image from each position in the reference set. This filter allows ambiguous image matches to be resolved.

### Related Work

Nelson [1991] proposed an early visual homing system for a mobile robot which captured images throughout the environment during a training phase. During the test phase,

the robot would find the image most similar to the current image which would have an associated movement direction that the robot would follow to return home. In this case, the objective was not to compute a localization solution but to move directly to a pre-defined goal position. This approach is similar to a biological model for honeybee visual homing [Cartwright and Collett, 1987].

There has been significant work on the best way of representing high-dimensional imagery data using some low-dimensional signature which represents the whole image. This vein of research operates on omnidirectional or panoramic images which can yield rotationally-invariant descriptions of the current position. Ulrich and Nourbakhsh [2000] represented positions by extracting image histograms from omnidirectional images. Lamon et al. [2001] encoded the sequence of vertical edges and coloured regions in an omnidirectional image into a compact string of characters. Comparisons between strings could be made quite efficiently by adopting fast string matching methods. Kröse et al. [2001] used principal components analysis to project new images into a lower-dimensional space for comparison. Menegatti et al. [2004] applied the Fourier transform on the rows of an omnidirectional image and used the lower-frequency components of the magnitude spectra as a low-dimensional signal.

Another approach to image-based localization is based on the extraction of SIFT keypoints and other related means of extracting and characterizing image sub-regions in a scale, orientation, and illumination invariant manner.

Se et al. [2001] presented a localization approach based on SIFT keypoints whose three-dimensional positions in space are estimated using stereo vision, with position uncertainty represented and managed via a Kalman filter. A recent trend is to extract keypoints and associated descriptors, then cluster the descriptors to form a set of representative “words” so that analogous methods to text retrieval can be used for fast classification of new “documents” (i.e., new images) [Sivic and Zisserman, 2003]. Filliat [2008] applied this approach to localize a robot with the visual vocabulary being constructed incrementally. Cummins and Newman [2008] have taken this approach even further by developing a model that assigns a low probability to observations that are indistinct and therefore unhelpful in resolving the robot’s current position. Image-based localization has also been employed as a component of various navigation systems, particularly for route following [Zhang and Kleeman, 2009; Chen and Bircheld, 2009; Furgale and Barfoot, 2010; Vardy, 2010].

AUV navigation techniques have traditionally relied upon acoustic transponders for localization. These transponders consist of at least two transducers and are typically classified by the distance between transponders. This distance is known as the baseline and the transponders are referred to as ultra-short baseline (USBL), short baseline (SBL), or long baseline (LBL) which implies multiple units that are either fixed relative to the seabed or floating on the surface with GPS access. However, most relevant to this work is the newer trend of localizing an AUV with respect to features of the environment.

We refer the reader to Paull et al. [2014] for a recent and comprehensive review of AUV localization and navigation which includes techniques based on acoustic transponders as well as those based on detecting features of the environment.

These features may be perceived through optical, sonar, or even magnetic sensors. There have been a number of interesting approaches to AUV navigation which are based on either single-beam or multibeam sonar [Meduna et al., 2010; Claus and Bachmayer, 2014]. These sonar technologies yield ranges whereas our work is based on side scan sonar which yields intensity-based images.

Side scan sonar images can often be interpreted similarly to optical images, yet they have very distinctive properties such as the nadir, the presence of acoustic shadows, and range-varying attenuation and resolution [Blondel, 2009]. Side scan sonar images have been used as part of a complete framework for Simultaneous Localization and Mapping (SLAM) as early as 2004 [Ruiz et al., 2004], although detection of natural landmarks from

side scan images was not directly considered. Other papers that have incorporated side scan imagery for SLAM include Pinto et al. [2009] and Woock and Frey [2010]. In terms of work that has focused on extracting landmarks (a.k.a. features) from side scan sonar images, we are aware of only a few other attempts.

Stalder et al. [2008] propose a landmark detection scheme work that is based on disruptions in the textures observed within side scan images. Aulinas et al. [2011] consider bright spots with associated dark shadows as potential landmark components but add robustness by wrapping this approach within the Haar cascade framework. Padiyal et al. [2013] consider the acoustic shadows produced by an object protruding from the seabed and utilize a correlation measure between observed shadows and those predicted from bathymetry for terrain-relative navigation.

### System Overview

The proposed system uses raw side scan sonar data along with standard navigation data to generate a corrected position estimate. Figure 3 illustrates the processing pipeline. A new update

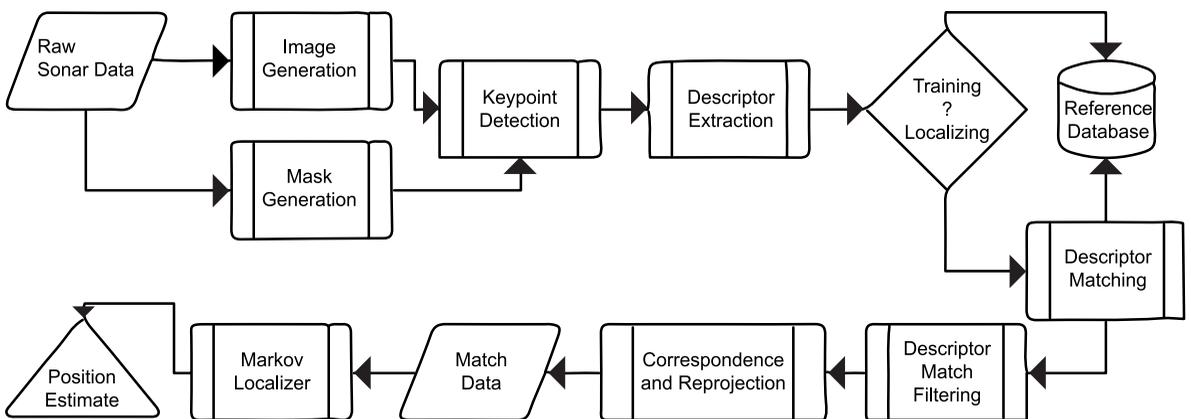


Figure 3: Processing pipeline diagram.

cycle is started when a threshold number of sonar pings have been recorded. The raw data is automatically processed into a geographically corrected image. During image generation some important navigation data such as the extents of the swath coverage and first and last positions are extracted. Keypoints are detected in the image and are stored along with the navigation information. Feature descriptors are then computed for each keypoint to be used for image matching. If the system is in the training phase, the feature descriptors, keypoints and navigation information are stored in the reference database. When the system is in the localizing phase, this information is sent directly to the matching process. The newly collected descriptors are compared to the descriptors collected in the training phase. The resultant descriptor matches are filtered based on the sizes and orientations of matched keypoints. In other words, if two keypoints are matched but differ strongly in size or orientation, then that match will be deleted. The surviving matches are then processed to determine the  $x,y$  translation that maps the query keypoints onto the reference image. This translation is treated as the motion input to the localizer.

Over each update cycle the localizer considers both the matched keypoints and the computed translation as it attempts to estimate the vehicle's position.

The particulars of image generation are not considered in this work. The focus is on image matching and its application to AUV localization. A description of the tools used to convert sonar data into images has been provided in King et al. [2014], King et al. [2012], and Vandrish et al. [2012].

## IMAGE MATCHING AND REGISTRATION

Image matching is the process of determining whether two images acquired at different times are of the same scene. A pair of images are considered a match if the registration process yields a good solution. During the registration process the parameters that describe the transformation of the query scene to the reference scene are computed. These parameters further refine the knowledge of position from node scale (tens or hundreds of metres) to metre scale.

Feature based image registration has become a mature technique in the last decade and several highly effective algorithms have been developed. Perhaps the most influential feature extraction technique is David Lowe's SIFT (Scale-Invariant Feature Transform) [Lowe, 2004]. SURF (Speeded Up Robust Features) [Bay et al., 2008] was developed by Herbert Bay et al. in 2006. SURF performs several times faster than SIFT and is claimed to be more robust against different image transformations. Many other algorithms exist and we completed a partial survey and performance comparison for sonar image

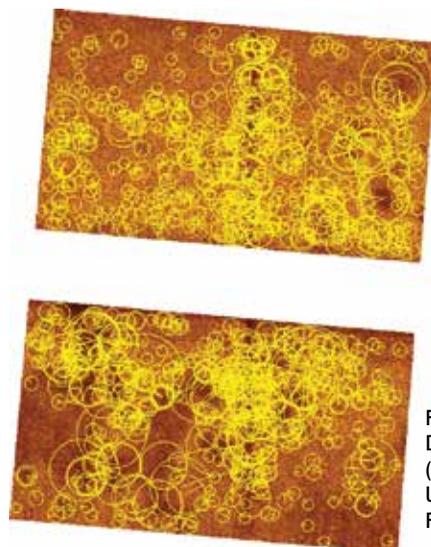


Figure 4: Detected SURF (Speeded Up Robust Features).

registration in King et al. [2013]. Figure 4 illustrates a sonar image and the corresponding extracted features.

Keypoints are detected in the image by finding local features that the particular keypoint extraction algorithm considers good for matching. Multiple algorithms can be employed at the same time to allow for detection of diverse feature types.

A keypoint contains a centre point where the feature is located; a radius to define the encompassing area; an angle representing the orientation of the feature; and a response which is a metric of feature strength. The set of keypoints for a particular image is limited to a maximum size of 1,000 to bound the computational complexity of descriptor extraction and matching. The first 1,000 keypoints with the highest response values are retained. The keypoints are then checked to see if their radius is within the bounds of the image mask. This rejects keypoints that may be generated from false features in noisy areas, turns, or from the image boundaries.

### Descriptor Extraction

For each feature keypoint a corresponding feature descriptor is computed. The descriptor is a vector of values that represent the feature in a way that is invariant to scale, orientation and other transformations. The descriptors are coupled with navigation and position information to form a *pathTile*. If the collection was made during the training phase, the *pathTile* is added to a reference database. If it was collected during the localization phase, the *pathTile* is immediately used to be matched against the reference database.

### Descriptor Matching and Filtering

During the localization phase the most recently collected *pathTile* descriptors are matched against the reference database collected during the training phase. A simple distance metric is computed for each pair of query and reference descriptors by using the  $L^2$  norm for SIFT and SURF descriptors. A smaller distance value correlates to a better match.

Since the sonar images are formed with north aligned up and are slant-range corrected, the difference in scale and angle between two views of the same feature should ideally be zero. Descriptor matches are further filtered by comparing the query and reference keypoints' sizes and angles. If the absolute difference between the size and angle of a query and reference keypoint exceeds a specified threshold, the match is rejected. This filtering is controlled by two threshold parameters: *sizeMatchThreshold* and *angleMatchThreshold*.

The keypoint extraction and matching algorithms allow a wide array of possible transformations, including both rotation and translations, between query and reference scenes. In our case, there should only be an  $x$  and  $y$  translation. By using algorithms that are meant for a broader class of transformations, we increase the system's robustness to perceived pose change due to inconsistent image generation caused by inaccurate altitude estimates and other environmental factors. Setting the *sizeMatchThreshold* and *angleMatchThreshold* to larger values allows some of this variation to be included. However, we have found that reducing these thresholds is also a very effective way of reducing the number of false positive matches.

To determine the optimal values for these thresholds, a binary classification performance analysis was performed over a range of values. The process is discussed in detail in “Binary Classification Performance Evaluation” later in this paper.

### Correspondence and Reprojection

To compute the 2D translation that transforms the query scene into the reference scene, a specialized correspondence algorithm was developed. In the case of general 3D pose estimation, an algorithm such as RANSAC [Fischler and Bolles, 1981] is typically used. RANSAC tests random subsets of correspondences to identify a subset that has significant agreement with the whole set of raw correspondences (measured as the number of inliers). This search for a good subset is bounded by an iteration limit.

In our case there should only be a 2D translation. Thus, a single correspondence can provide a translation that can be tested against the whole set of raw correspondences directly. In other words, the reduced space of possible transformations means that we can compute a translation for each correspondence, then count the number of inliers as the number of other correspondences that are in agreement.

The full algorithm is provided as Algorithm 1: *queryKpSet* and *refKpSet* are the query and reference keypoint sets; *matches* is

the set of descriptor matches computed during the matching phase. A *match* contains the index of the query and reference keypoints that generated the match. When the algorithm has finished, *mostInliers* is the number of inliers retained; *lowestError* is the total reprojection error; and *Delta\_x\_best*, *Delta\_y\_best* is the translation that describes the transformation from query to reference scene.

In brief this algorithm iterates through all the feature matches and determines how well the other matches support them. The match which results in the highest number of inliers (most support) is chosen to compute the translation. Figure 5 is an example of two images and the corresponding inlier matches.

---

```

1: function FINDCORRESPONDENCE(matches, queryKpSet, refKpSet, errorThresh)
2:    $\Delta x_{best} \leftarrow 0$ 
3:    $\Delta y_{best} \leftarrow 0$ 
4:    $lowestError \leftarrow \infty$ 
5:    $mostInliers \leftarrow 1$ 
6:   for  $i \leftarrow 0, matches.size()$  do ▷ Try each ref. keypoint as start seed
7:      $im \leftarrow matches[i]$ 
8:     ▷ Compute the delta between ref. and query keypoint locations
9:      $\{\Delta x, \Delta y\} \leftarrow refKpSet[im.refIdx].pt - queryKpSet[im.queryIdx].pt$ 
10:     $totalError \leftarrow 0$ 
11:     $numInliers \leftarrow 0$ 
12:     $kpErrors \leftarrow \emptyset$  ▷ List to store used keypoints
13:     $kpErrors[im.refIdx] \leftarrow 0.0$ 
14:    ▷ Reproject all query keypoints based on the seed deltas
15:    for  $j \leftarrow 0, matches.size()$  do
16:      if  $i = j$  then ▷ Error is 0 for the seed
17:        continue
18:      end if
19:       $jm \leftarrow matches[j]$ 
20:       $\{tx, ty\} \leftarrow queryKpSet[jm.queryIdx].pt + \{\Delta x, \Delta y\}$ 
21:       $error \leftarrow \|\{tx, ty\} - refKpSet[jm.refIdx].pt\|$ 
22:      if  $error < errorThresh$  then ▷ Check error for inlier status
23:        if  $kpErrors[jm.refIdx] = \emptyset$  then
24:           $numInliers \leftarrow numInliers + 1$ 
25:           $totalError \leftarrow totalError + error$ 
26:           $kpErrors[jm.refIdx] \leftarrow error$ 
27:        else if  $error < kpErrors[jm.refIdx]$  then
28:           $totalError \leftarrow totalError - kpErrors[jm.refIdx] + error$ 
29:           $kpErrors[jm.refIdx] \leftarrow error$ 
30:        end if
31:      end if
32:    end for
33:    ▷ Check if this seed produced a better result
34:    if  $numInliers > mostInliers$  ( $numInliers = mostInliers \& totalError < lowestError$ )
35:       $mostInliers \leftarrow numInliers$ 
36:       $lowestError \leftarrow totalError$ 
37:       $\Delta x_{best} \leftarrow \Delta x$ 
38:       $\Delta y_{best} \leftarrow \Delta y$ 
39:    end if
40:  end for
41:  return  $mostInliers, lowestError, \Delta x_{best}, \Delta y_{best}$ 
42: end function

```

---

Algorithm 1: Correspondence algorithm

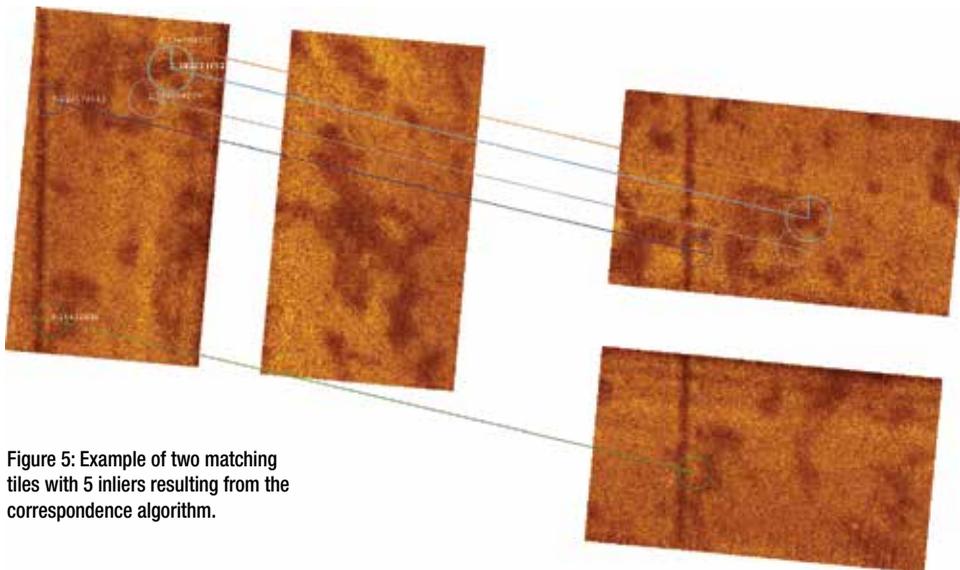


Figure 5: Example of two matching tiles with 5 inliers resulting from the correspondence algorithm.

### Definition of Parameters

Parameters for sonar image registration can be generalized into two categories. Image generation parameters affect how the raw sonar data is transformed into image space thus indirectly affecting registration performance. Image registration parameters directly affect the registration algorithm.

**gridResolution** The size (in metres) an image pixel represents in units of m/pixel. Higher resolution imagery increases quality of registration but also increases image size and computation time. For maximum quality, the resolution should be comparable to the maximum resolution of the sonar system used for collection, although if the survey speed is too high, using a lower resolution will help close gaps in along track coverage. In this work, the resolution is 0.2 m/pixel.

**maxSwathWidth** The maximum distance from the nadir to either side of the swath that is considered to be of good quality. At higher altitudes from the sea floor the data closer to the outer edge of the swath becomes noisy. This adversely affects registration.

**swathBlanking** The distance from the nadir to either side of the swath that is considered poor quality.

**distinctMatchThreshold** During the process of descriptor matching, possible match candidates are computed. The “distinctMatchThreshold” specifies the minimum difference in distance between the best match candidate and the next best candidate. When the difference is at least greater than the specified threshold, the match is considered distinct and the match is retained. Otherwise the match is rejected.

**sizeMatchThreshold** The maximum absolute difference in size between a keypoint match pair for the pair to be retained. When sonar images are georeferenced with north facing up and corrected for altitude, the detected features should have comparable scale and orientation. Due to changes in viewpoint, however, they may not be exactly equal.

**angleMatchThreshold** The maximum absolute difference in angle between a keypoint match pair. Again, for the same

reasoning as above, the same features between two images should have comparable angles.

**reprojThreshold** To judge the strength of a putative correspondence, keypoint matches from one image are reprojected onto the corresponding positions of the other image. An individual keypoint's reprojection error is the distance between the keypoint and its reprojected position. For a set of matches, the root mean square (RMS) reprojection error is computed.

This threshold specifies the maximum RMS reprojection error that can be accepted for a particular correspondence between keypoints.

### Pre-Registration Image Preparation

Side scan sonar images have some inherent noise especially in the near nadir region and

towards the outer limits of the swath. Figure 6 illustrates the types of noise that may exist. Returns from the nadir region, directly beneath the vehicle's track, are weak and poorly sampled. This effect is indicated by B.

As the acoustic pings travel in the across track direction, the sound is attenuated. This is compensated by applying a time-varying gain. If the vehicle is at a high enough altitude such that the slant range in the outer swath is approaching the sonar's optimal range for a particular environment and configuration, the sound can be attenuated enough that these regions are poorly sampled and noisy. This effect is indicated by A in Figure 6.

If these areas of noise are present in every image, similar false features may be detected in all images leading to a high number of false

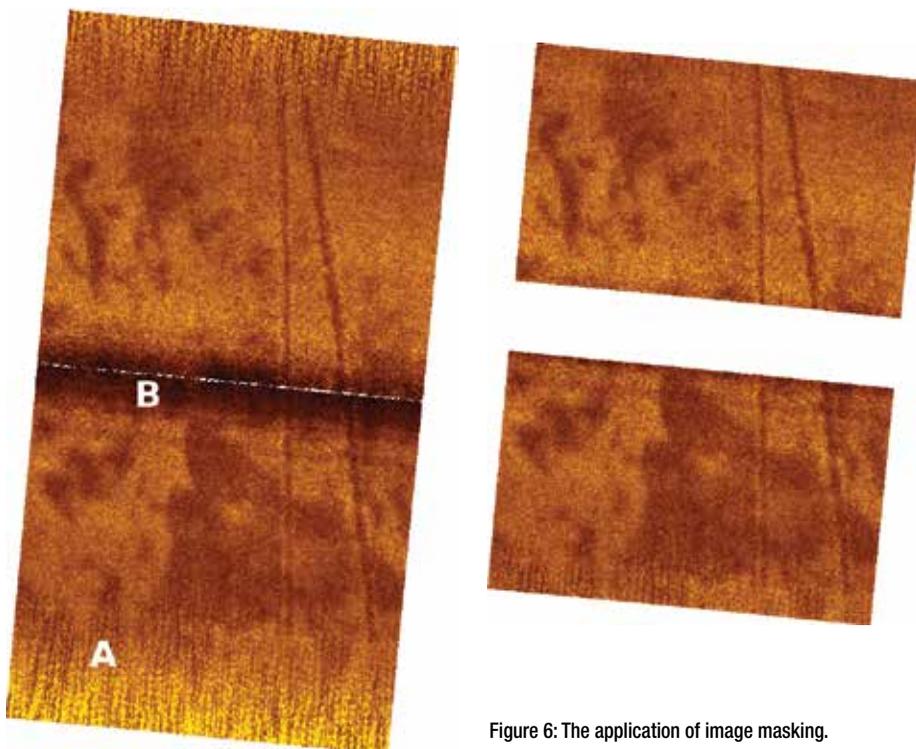


Figure 6: The application of image masking.

positive matches. To alleviate this problem, a mask is generated to remove these noisy areas and only expose areas that are of high quality thus reducing the number of detected false features. The mask is specified by the inner and outer limits of swath. Ping samples are plotted only if the sample's range is within these limits.

### **Binary Classification Performance Evaluation**

To measure the performance of the matching process and to quantify the effect of filtering descriptor matches by comparing their size and angle, a binary classification performance evaluation was performed on the image matching algorithm. The number of inlying keypoints determined by the correspondence algorithm measures the quality of a match.

Since this metric is a continuous value, it cannot be used directly for binary classification evaluation. Test interpretation is easier if the outcome is binary. If a cut-off value is applied to the number of inliers, the system is then converted into a binary classifier. In this case, a cut-off value of 4 inliers was chosen. A match with 4 or more inliers is considered a positive match.

There are four possible outcomes of a test. They are true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN). A true positive occurs when the classifier correctly identifies a match. A true negative occurs when the classifier correctly identifies a non-match. A false positive occurs when the classifier incorrectly identifies a match. A false negative occurs when the classifier fails to identify a match. Many

metrics can be formed from these four outcomes to measure the performance of a classifier [Vihinen, 2012]. Matthews correlation coefficient (MCC) is a balanced and non-biased measure of a classifier's performance. A MCC value ranges from -1 to 1. A value of 1 represents a perfect prediction; a value of 0 is no better than random prediction; and a value of -1 represents a complete disagreement between prediction and observation. The MCC score is calculated as follows:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

### **LOCALIZATION**

Given a previously collected set of sonar tiles covering a defined geographic area, an AUV that is collecting sonar tiles can localize itself using only the results of image registration. Localization can take the form of true global localization, in which the AUV has no initial knowledge of its position, and tracking, where an initial estimate is known and the localization improves or prevents the estimate from degrading over time [Thrun et al., 2005].

This section describes a grid-based Markov localization implementation. The belief at each update step is based on the previous belief, knowledge about the AUV's motion from the previous step, and information about how the AUV's current observation matches against the previously collected reference set. In this work we only determine the absolute position in latitude and longitude, with no determination of pose. Thus the belief grid is limited to two dimensions.

## Implementation

Localization is implemented in Python using field collected data and matching information provided by the results of image registration. Following initialization of the belief grid, a straightforward implementation of Bayesian filtering is applied in which a predictive step shifts the belief based on knowledge about the AUV's movement. An observation is then made by matching the current sonar data image to the reference set. The belief grid representing the probability of making this observation is then multiplied by the predicted belief and the result is then normalized. After each update a single estimate of position is inferred from the belief grid.

### *Initialization*

A two-dimensional matrix is created in memory to hold the belief value of each potential location. Belief values are between 0.0 and 1.0, indicating no belief to full belief, respectively. The tiles representing the reference set are parsed to determine the maximum and minimum latitudes and longitudes that can be localized. The size of the grid is then set based on the resolution, in this case  $\frac{1}{1000}$  of a degree in each dimension. This gives a geographic resolution of approximately 11 m in latitude and 9 m in longitude. This resolution is sufficient as the average tile size is approximately 100 m<sup>2</sup>.

For the global localization problem, the initial belief values are normalized across all locations such that the total belief for all locations is 1.0. This is accomplished by setting each location value to:

$$\frac{1.0}{(xDim) (yDim)}$$

For the tracking problem, an assumed knowledge of initial position is used to seed the belief grid. In this case the location is assumed known for the first observation. The belief values for the area covered by the observation tile are set to a value of twice the mean value of the un-seeded normalized grid. The belief grid is then normalized again to get a total belief of 1.0.

### *Prediction*

To perform the prediction aspect of Markov localization, it is important to shift the belief between observations to account for the vehicle's movement. The sonar record contains information about the AUV's internal estimate of position, its heading and speed. The delta position for an update cycle is taken as the difference between the last capture positions of the current and previous tiles. As this is strictly a position difference, any errors in the absolute measurements are ignored, thus separating the localization scheme estimate from the AUV's own estimate. The delta position is then applied to the belief map by performing a shift operation in each dimension. A Gaussian blur is then applied to account for errors in the motion estimate.

As no estimate of the position change can be perfect, potential errors must be accounted for. This error depends on the system being used and can range from 0.1% of distance travelled in a high-end system to 5% in a lower end system.

### *Observation and Update*

As the AUV obtains a new sonar tile, the Image Registration system will match the tile

against the reference set. The results of the Image Generation are presented as a set of matching information for each element in the reference set, including the number of inliers for each match case and the total reprojection error. Results from previous testing have shown that the number of inliers is the dominant indicator of a strong match, with numbers less than 2 indicating a non-match. With this in mind, the weight of a match is 0 for inlier counts of less than 2, otherwise the weight of the match is calculated from the following equation:

$$weight = 0.5 \frac{\min(1, num\_inliers^2)}{total\_Error}$$

Where weight is zero when inlier count is less than 2, the number of inliers provides a squared component which is then divided by the total reprojection error. The constant 0.5 is a gain parameter applied to reduce the overall effect of all measurements. As the Image Generation system tends to provide a large number of false-negatives and false-positives, the weight is reduced and the base value for all measurement locations is increased. For these results, a measurement grid is created and normalized so that the sum of all locations is 0.5. The value of 0.5 comes from tests run on the collected data and should not be considered optimal. The areas covered by a match are increased by the weight value.

The measurement grid is then multiplied by the belief grid to perform the update step.

#### *Normalization and Estimate*

The resultant belief map is normalized

such that the sum of all belief locations is 1.0. To track the performance of the localizer, we need to extract a single estimate from this step which is recorded and compared against knowledge about the actual AUV position.

A determination of the centroid, or first grid moment, is taken as an estimate of the weighted average position. This value is then recorded along with the grid area as the current estimate of each update step. As there is a period of convergence before an estimate can be taken as valid, the entropy of the belief grid is calculated and used to determine the validity of the estimate. Prior knowledge of the system and off-line tests are used to develop an entropy threshold at which the estimate is acceptable.

## RESULTS

Testing was performed on a survey data set collected by Memorial University's Explorer AUV during trials in Holyrood Bay, Newfoundland, Canada, as shown in Figure 7. The survey consists of 25 parallel lines running east-west and three transect lines running north-south. Two of the north-south lines are used as the reference set and the east-west lines are used as the query set. The raw files are split into equal sized tiles consisting of 1,000 sonar pings.

### **Image Matching**

Matching was performed with values of *sizeMatchThreshold* and *angleMatchThreshold* ranging from 5 to 60 in steps of 5 which results in a total of 144 test cases. For each case, the number of false



Figure 7: Surveyed area used for testing.

positives, true positives, false negatives and true negatives were determined using the area of coverage overlap between tiles. This dataset was chosen in part because the query set was collected orthogonal to the reference set thus making the test more challenging due to the difference in ensonification perspectives.

Figures 8 and 9 show the true positive and false positive counts for varying threshold values. Increasing both angle and size thresholds results in a higher number of false positives. It also increases the number of true positives but the effect diminishes as the threshold values approach their maximum values. Figure 10 is a plot of the MCC score against size and angle thresholds. The MCC score is meant to quantify the trade-off between false and true positive matches. The plot correlates nicely with Figures 8 and 9

and it can be seen that an optimal configuration in terms of MCC is in the location of the peak. The peak value of the MCC score is 0.3343 with a size threshold of 15 and an angle threshold of 20. Figure 11 shows the average localizer error plotted against size and angle thresholds. It can be seen that the region of lowest error correlates to the region of the highest MCC score.

### Localization

For each test, both global and tracking localization was performed. Results from localization were compared against the vehicle's position estimate recorded during collection.

As matching results were computed for a variety of input parameters, localization results are presented at three intervals over the parameter space.

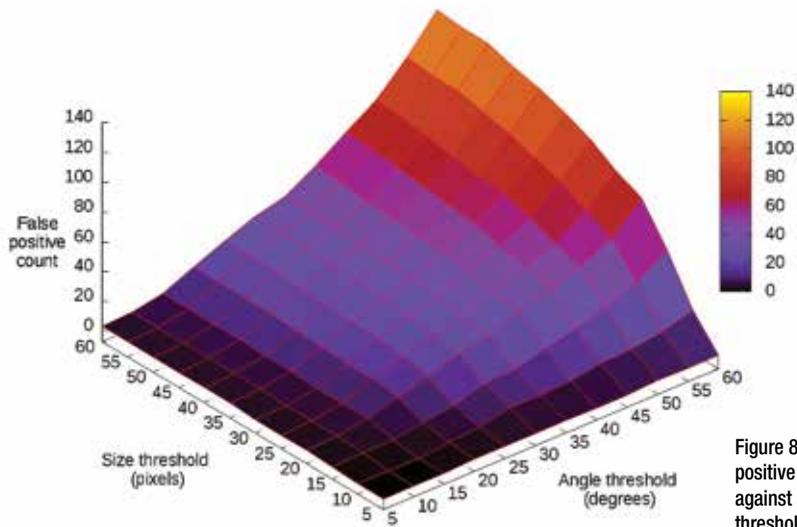


Figure 8: False positive count plotted against size and angle thresholds.

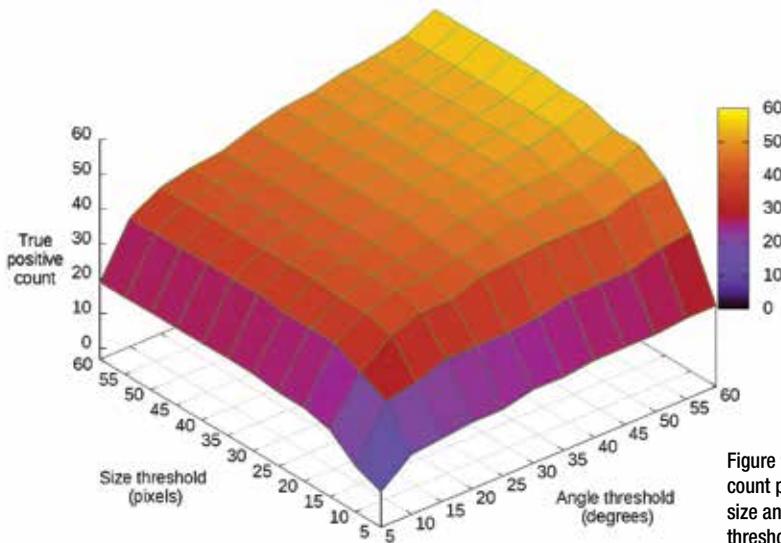


Figure 9: True positive count plotted against size and angle thresholds.

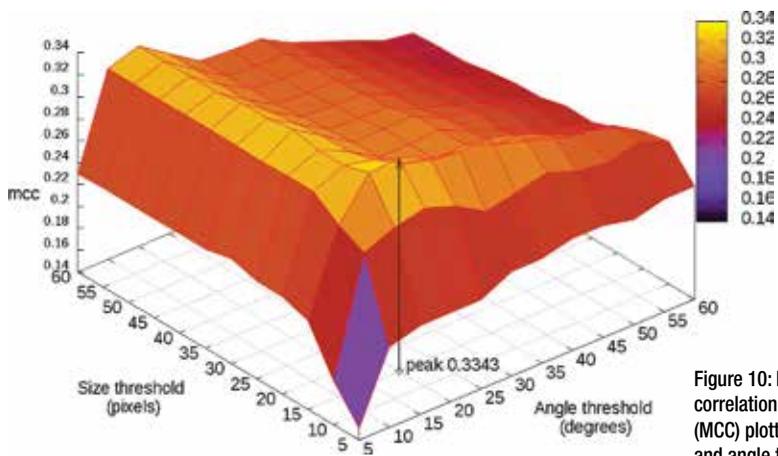


Figure 10: Matthews correlation coefficient (MCC) plotted against size and angle thresholds.

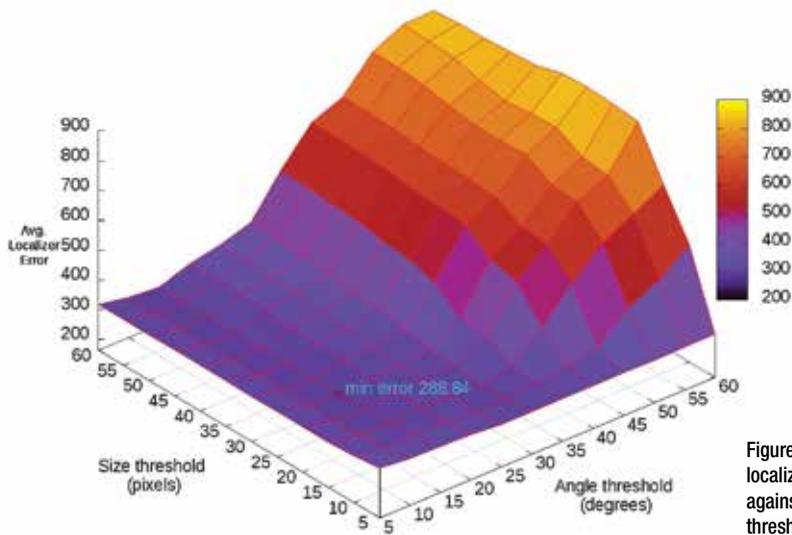


Figure 11: Average localizer error plotted against size and angle thresholds.

We will show localization performance at angle-size threshold pairs of (10,10), (30,30), (60,60).

### *Global Localization*

For global localization, the initial belief grid is normalized to a uniform distribution. At each observation step, the match information is parsed and input into the Bayesian filter as a measurement. After each step, the centroid of the belief is used to determine the peak belief and this is considered our estimate of location. At this time, the entropy of the belief map is determined as a metric of how good this estimate is. The absolute difference between the estimate and the actual observation position is also calculated as a measure of the positional error at the current step.

Figures 12, 13, and 14 show the results of the localizer over three parameter settings. In each graph, the position error is plotted as a line, with the entropy value as vertical bars. A horizontal threshold line is added at 100 m, which corresponds to the average side length of a reference tile. We consider error values below this line to indicate successful localization: that is, localization to the correct

image from the database has been achieved if the error is below 100 m.

For the (10,10) and (30,30) cases, we see that the initial position error is high. This is expected since we start with no prior knowledge and our initial estimate is arbitrary. As observations are made, we see the error decrease. We also notice the entropy value decrease as we make updates and increase our belief confidence. In both cases we see the error eventually go below the error threshold and the entropy decrease to the point that we can call the localization successful.

For the (10,10) case, the entropy does not initially decline as fast as the (30,30) case; this is due to a reduced number of observations with tighter filtering.

The (60,60) case, in which filtering is very liberal allowing many false and true positive matches, does not achieve a successful localization in terms of position error, but we do see the entropy fall off quickly. This indicates that our localizer has made a confident estimate of position, but it is not correct – it makes a false positive localization.

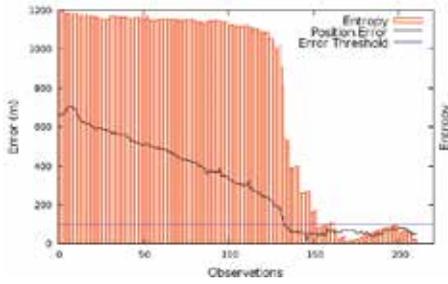


Figure 12: Global localization for angle 10, size 10.

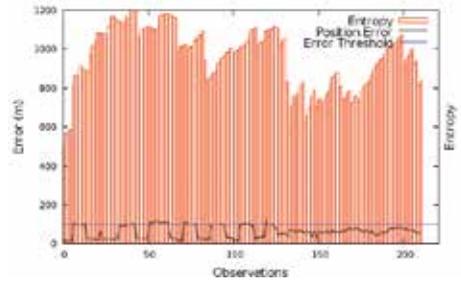


Figure 16: Tracking localization for angle 10, size 10.

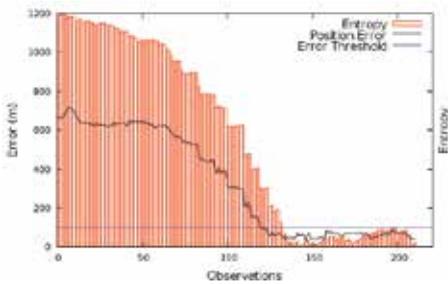


Figure 13: Global localization for angle 30, size 30.

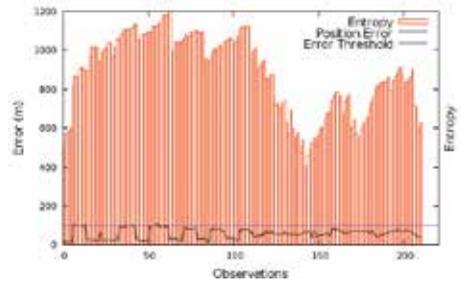


Figure 17: Tracking localization for angle 30, size 30.

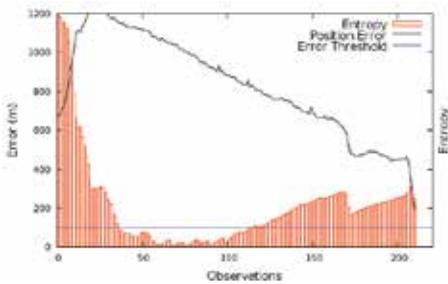


Figure 14: Global localization for angle 60, size 60.

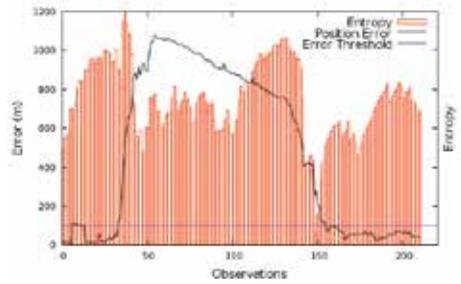


Figure 18: Tracking localization for angle 60, size 60.

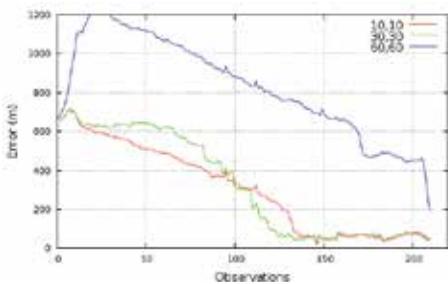


Figure 15: Combined global localization results.

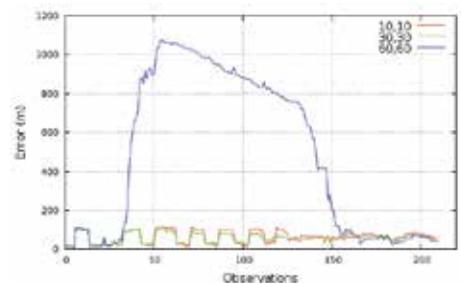


Figure 19: Combined tracking localization results.

As more observations are made, we eventually begin to converge on the correct solution, but not before the test ends.

Figure 15 plots the performance of all three parameter sets. We see that (30,30) is the fastest to converge, with (10,10) reaching a solution as well. (60,60) does not converge and initially provides a worse error.

From these results, we see that localization is possible with this scheme, but depends on the tuning of the measurement system – Image Registration. Conservative filtering limits the number of observations, while liberal filtering allows more matches, but can be considered noisy with both good and bad matches.

Though the resolution is limited to reference set size, the localizer does work and can provide a position estimate given no prior information.

### *Tracking*

The tracking localizer is identical to the global localizer except that the initial position is seeded with the first observation location. Thus it starts off with a solution. As in the previous section, we plot performance for (10,10), (30,30), (60,60) values of angle and size thresholds. Figures 16, 17, and 18 show the results of the localizer over three parameter settings.

Again the (10,10) and (30,30) cases both successfully maintain their estimate below the threshold – with the (30,30) case providing slightly better error performance in the middle region. (60,60) again diverges to a false estimate even while the entropy remains low.

Figure 19 shows the comparative performance of each parameter set.

## CONCLUSION

Without the luxury of global positioning satellites, long term autonomy in the depths of the ocean is a much more challenging feat than for terrestrial robotics. This paper proposed a solution to this problem based on revisiting previously surveyed areas and comparing the current side scan sonar image to a reference set. It has been shown that as long as the target area is not completely featureless, our technique can be successful. It has been demonstrated that even using a very challenging dataset where the vehicle tracks orthogonal to the reference track a position estimate will converge to the real global position. It has also been shown that the sensitivity of image matching can be controlled by adjusting the size and angle thresholds for keypoint match filtering.

One potential issue with this kind of navigation is that the seabed is ever-changing. Tides and currents can alter the appearance of the seabed over time. This is not a problem for areas with persistent features like rock outcroppings. Another potential issue could be caused by long linear features that appear the same along their path. This is not a problem if there are sufficient distinctive features amid the positionally ambiguous linear features. In our experiments, spurious mismatches are easily filtered out and do not significantly impair our localizer.

### **Future Work**

Future work will include in-field testing of this

system. Ideally the system could be tested over various seabed types and in hostile environments such as under sea ice. Various reference path patterns could be explored to determine the optimal size and spacing of pre-surveyed regions to maintain stable localization. For example, if returning from a long deployment where the angular error is expected to grow substantially, the pre-surveyed region might be elongated to ensure that the vehicle's likely return trajectories are covered. The pre-surveyed area might also be extended if the imagery data is believed to be indistinct and therefore not useful for localization. Path searching algorithms could be developed to enable the vehicle to locate the pre-surveyed area in a search pattern based on the current error estimate.

## REFERENCES

- Alahi, A.; Ortiz, R.; and Vandergheynst, P. [2012]. *FREAK: Fast retina keypoint*. In proceedings of IEEE Computer Vision and Pattern Recognition (CVPR), pp. 510-517.
- Aulinas, J.; Fazlollahi, A.; Salvi, J.; LLado, X.; and Garcia, R. [2011]. *Robust automatic landmark detection for underwater SLAM using side-scan sonar imaging*. In proceedings of International Conference on Mobile Robots and Competitions, Lisboa (Portugal), p. 70.
- Bay, H.; Ess, A; Tuytelaars, T.; and Gool, L.V. [2008]. *SURF: Speeded up robust features*. Computer Vision and Image Understanding (CVIU).
- Blondel, P. [2009]. *The handbook of sidescan sonar*, Springer Praxis.
- Cartwright, B. and Collett, T. [1987]. *Landmark maps for honeybees*. Biological Cybernetics, Vol. 57, pp. 85-93.
- Chen, Z. and Bircheld, S. [2009]. *Qualitative vision-based path following*. IEEE Transactions on Robotics, Vol. 25. No. 3, pp. 749-754.
- Claus, B. and Bachmayer, R. [2014]. *Towards online terrain aided navigation of underwater gliders*. In proceedings of IEEE/OES Autonomous Underwater Vehicles.
- Cummins, M. and Newman, P. [2008]. *FAB-MAP: Probabilistic localization and mapping in the space of Appearance*. International Journal of Robotics Research, Vol. 27, No. 6, pp. 647-665.
- Filliat, D. [2008]. *Interactive learning of visual topological navigation*. In proceedings of IEEE/RSJ International Conference on Robots and Systems (IROS).
- Fischler, M.A. and Bolles, R.C. [1981]. *Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography*. Communications of the ACM, Vol. 24, No. 6, pp. 381-395.
- Furgale, P. and Barfoot, T. [2010]. *Visual teach and repeat for long-range rover autonomy*. Journal of Field Robotics, Vol. 27, No. 5, pp. 534-560.
- King, P; Vardy, A; Anstey, B.; and Vandrish, P. [2012]. *Real-time image generation and registration framework for auv route following*. In proceedings of IEEE-AUV.
- King, P.; Anstey, B.; and Vardy, A. [2013]. *Comparison of feature detection techniques for AUV navigation along a trained route*. In proceedings of MTS/IEEE OCEANS, pp. 1-8.

- King, P.; Anstey, B.; and Vardy, A. [2014]. *Preliminary field trials of autonomous path following*. In proceedings of IEEE-AUV.
- Kröse, B.J.; Vlassis, N.; Bunschoten, R.; and Motomura, Y. [2001]. *A probabilistic model for appearance-based robot localization*. Image and Vision Computing, Vol. 19, No. 6, pp. 381-391.
- Lamon, P.; Nourbakhsh, I.; Jensen, B.; and Siegwart, R. [2001]. *Deriving and matching image fingerprint sequences for mobile robot localization*. In proceedings of ICRA, pp. 1609-1614.
- Lowe, D. [2004]. *Distinctive image features from scale-invariant keypoints*. International Journal of Computer Vision, Vol.60, No. 2, pp. 91-110.
- Meduna, D.K.; Rock, S.M.; and McEwen, R.S. [2010]. *Closed-loop terrain relative navigation for AUVs with non-inertial grade navigation sensors*. In proceedings of IEEE/OES Conference on Autonomous Underwater Vehicles, pp. 1-8.
- Menegatti, E.; Maeda, T.; and Ishiguro, H. [2004]. *Image-based memory for robot navigation using properties of omnidirectional images*. Robotics and Autonomous Systems, Vol. 47, No. 4, pp. 251-267.
- Nelson, R. [1991]. *Visual homing using an associative memory*. Biological Cybernetics, Vol. 65, No. 4.
- Padial, J.; Dektor, S.G.; and Rock, S.M. [2013]. *Correlation of sidescan sonar acoustic shadows and bathymetry for terrain-relative navigation*. Unmanned Untethered Submersible Technology.
- Paull, L.; Saeedi, S.; Seto, M.; and Li, H. [2014]. *AUV navigation and localization: A review*. Journal of Oceanic Engineering, Vol. 39, No. 1, pp. 131-149.
- Pinto, M.; Ferreira, B.; Matos, A.; and Cruz, N. [2009]. *Using side scan sonar for relative navigation*. In proceedings of Industrial Electronics.
- Ruiz, I.T.; De Raucourt, S.; Petillot, Y.; and Lane, D.M. [2004]. *Concurrent mapping and localization using sidescan sonar*. Journal of Oceanic Engineering, Vol. 29, No. 2, pp. 442-456.
- Se, S.; Lowe, D.; and Little, J. [2001]. *Mobile robot localization and mapping with uncertainty using scale invariant visual landmarks*. The International Journal of Robotics Research, Vol. 21, No. 8, pp. 735-758.
- Sivic, J. and Zisserman, A. [2003]. *A text retrieval approach to object matching in videos*. In proceedings of IEEE Computer Vision, pp. 1470-1477.
- Stadler, S.; Bleuler, H.; and Ura, T. [2008]. *Terrain-based navigation for underwater vehicles using sidescan sonar images*. In proceedings of MTS/IEEE Conference on Oceans.
- Thrun, S.; Burgard, W.; and Fox, D. [2005]. *Probabilistic robotics*. MIT Press, Cambridge, MA.
- Ulrich, I. and Nourbakhsh, I. [2000]. *Appearance-based place recognition for topological localization*. In proceedings of IEEE ICRA, Vol. 2, pp. 1023-1029.
- Vandrish, P.; Vardy, A.; and King, P. [2012]. *Towards AUV route following using qualitative navigation*. In proceedings of IEEE Canadian Conference on Computer and Robot Vision.
- Vardy, A. [2010]. *Using feature scale change for robot localization along a route*. In proceedings of IEEE/RSJ International

Conference on Robots and Systems.

Vihinen, M. [2012]. *How to evaluate performance of prediction methods measures and their interpretation in variation effect analysis*. BMC Genomics, Vol. 13, Supplement 4.

Woock, P. and Frey, C. [2010]. *Deep-sea AUV navigation using side-scan sonar images and slam*. In proceedings of IEEE OCEANS, pp. 1-8.

Zhang, A. and Kleeman, L. [2009]. *Robust appearance based visual route following for navigation in largescale outdoor environments*. The International Journal of Robotics Research, Vol. 28, No. 3, pp. 331-356.