Towards Three-Dimensional Heterogeneous Imaging Sensor Correspondence and Registration for Visualization

Matthew Johnson-Roberson, Oscar Pizarro, and Stefan Williams
ARC Centre of Excellence for Autonomous Systems
Australian Centre for Field Robotics, The University of Sydney
{m.roberson, o.pizarro, stefanw}@acfr.usyd.edu.au

Abstract—The registration and 3D mosaicing of range images presents an interesting challenge for the underwater domain. This paper attempts to lay out some of the issues involved in generating large 3D reconstructions of the benthos from two distinct views gathered from an AUV. Specifically, we address using 3D feature correspondence to estimate a Euclidean transformation between views and then refine that estimate to generate visually consistent meshes for human viewing. To achieve this goal we implement spin image feature matching in a pipeline that robustly processes range points to output registered meshes.

I. INTRODUCTION

Seabed imaging is a crucial tool in evaluating and surveying large benthic environments. While several sensor technologies exist for subsea imaging, they all make tradeoffs in range vs. resolution and, even when multiple sensors are used, very little is done to fuse mapping/imaging sensors or even to guarantee their consistency. Registration also allows for the synthesis of data from multiple independent platforms surveying the same region without direct pose communication or full navigation coordination. This is becoming increasingly relevant with the great deal of attention being spent on coordination of multiple underwater vehicles the ability to fuse data from multiple perspectives and modalities [1]. Matching of 3D features has previously been addressed for several purposes, it has matched multibeam features with DEMs for navigation [2], and studied 3D mesh generation and 3D mosaicing from underwater imaging [3], [4]. We expand on this work through the exploration of 3D mesh registration in the context of multiple views and their integration into a unified mesh.

We are working towards a hierarchical approach of utilizing different views of the same location at multiple ranges and in future with various sensing modalities. The proposed technique is based on three-dimensional feature correspondence and registration. We generate and match these features to define rigid body transforms between point sets. This hierarchical approach takes range images at various resolutions and creates a large scale reconstruction from multiple 3D views. The application of such reconstructions are wide ranging from scientific use in environmental survey and underwater archeology to commercial use in drilling and pipe-laying. In this paper we discuss challenges of the problem in Section II followed by a discussion of our proposed technique in Section III and feature registration in depth in section III-B. Experiments in Section IV and finally we conclude and talk about future directions in Section V.

II. SETUP AND PROBLEM

A. Setup

There is a large body of work exploring sensor registration and 3D mesh creation [5]. The field of underwater 3D imaging is advancing quickly. New sonar technologies, positioning, and attitude sensing have allowed multibeam acoustic sensors to generate 3D maps of relatively high quality and resolution at ranges O(10-100m) [6]. Additionally, optical sensing such as stereo imaging can generate very high resolution maps, though at short ranges O(1m) and requiring clear water, at a fraction of the cost of high-end imaging sonars. Optical imaging also provides the capability of adding color image information to 3D reconstructions which is a promising new direction in benthic monitoring [7]. Despite high resolution imaging, navigation solutions are much less advanced and often there are large errors in pose estimates which prevent the use of data in its gathered form. Advances in imaging and errors in pose warrant finding ways of registering generic 3D data so that it may be at the resolution of the sensors rather than the navigation.

When compared to terrestrial mesh registration tasks, the subsea environment presents several interesting challenges. Terrestrial surveys can rely upon the aid of GPS and other high accuracy localization techniques which are not always available for underwater range imaging. Solely inertial and DVL navigation pose estimates have different errors than GPS based navigation solutions. Our proposed technique relies upon feature matching. In terrestrial tasks, many prominent unique geometric features exist which allow for relatively easy matching. The benthos is a difficult domain for geometric matching, as bottom features are often much less unique and harder to discriminate from one another [8].
B. Problem Statement

The basic goal of registering range images is to determine the Euclidean motion between two range views. We define one set of 3D points to be the reference mesh. This is the base reference frame into which we will transform the other points which we will refer to as the matching mesh. In the literature of range image registration this task is divided into two different types of registration which we will refer to as Coarse Registration and Fine Registration [9]. Coarse registration is used when no estimate of relative poses exists for your individual range images. Fine registration is the process of refining an existing estimate to improve its consistency and accuracy.

III. Method

We propose matching individual views with a reference mesh. In the following sections we describe a general overview of the approach adopted in this paper.

The data processing pipeline is outlined in Figure 1. Two point sets are turned into meshes for registration. The first mesh is defined as the reference mesh, the second being the matching mesh. To register these meshes we will generate features from each using spin images [10], described in depth in section III-B. After feature generation we match the two sets of features to find those with the highest correlation. We use Horn's algorithm [11] to determine the rigid body transformation relating corresponding points. Given that it is possible to propose incorrect correspondences based on correlation alone, we robustly determine the rigid transformation relating the two sets using RANSAC. Once the two views have been registered, they are processed to smooth and stitch the seams between the different sets of data.

A. Mesh Creation

The pipeline begins with mesh creation. Points from an appropriate sensor, such as multibeam sonar, are projected into space based on the current estimate of vehicle pose. A regular mesh is then generated based on a Delaunay triangulation of the resulting point cloud generated during successive passes over the terrain. Accurate pose estimates are important to the quality of the resulting mesh and the ability of the algorithm to find a consistent registration. The integration of DVL and compass, such as are available on many low cost Autonomous Underwater Vehicles, will result in locally consistent meshes but the estimate will tend to drift over time. Our aim is to provide a mechanism for registering these locally consistent meshes generated at different times or vantage points during a survey.

B. Coarse Registration

The range data under discussion can be gathered separately from multiple platforms and there is little guarantee that the views will all be orthogonal to the seafloor. The platforms have six degrees of freedom (6-DOF) in the water and for this reason, 3-DOF (x,y,z) registration techniques were not applicable to this work [11]-[13]. A 6-DOF resolution invariant matching technique was required for registration of these types of data sources [10].

Spin images provide an excellent characterization of geometric features in a point cloud. These features are robust and can be resolution invariant if calculated correctly [14]. A Delaunay triangulation is performed on the point set and normal vectors and tangent planes are calculated. Then any point in the cloud can be described by a spin image using two values: $\alpha$ the distance between normal vector at the target point and a point in the range image and $\beta$ which is the distance from the point in the range image to the tangent plane of the target point.

\[
\alpha = \sqrt{||x-p||^2 - (n(x-p))^2}
\]  

\[
\beta = n(x-p)
\]

$p$ is the 3D point, $n$ is the normal vector at the target point for the spin image, $x$ is a point around $p$.

After all the spinmap 3D features are generated in the reference mesh, the same feature computation is performed.
on the matching mesh to be registered. The subset of self-similar features for each mesh are determined and removed by correlating each of the spinnmaps with all of the other spinnmaps in that mesh. If they are similar above an empirically derived threshold, they are removed as they are unlikely to be reliably matched with similar features in the other mesh. A search is then performed to find features of highest correspondence. The remaining matched features are selected to be tested for validity. If additional information is available, for example relative poses, the performance can be improved by constraining the search for feature matches based on pose data from the platform’s position estimates. These matched features are filtered for false correspondences using the RANSAC method [16]. Once a set of consistent correspondences are obtained these features are used to generate a rigid-body transform between the two views.

C. Fine Registration

Once a coarse registration is established as described above, the two meshes have their registration iteratively refined with Iterative Closest Point (ICP) [17]. The distance between the closest point on another shape is iteratively minimized to refine the registration of two 3D shapes. After ICP is used to stitch the two meshes together a re-triangulation is performed on the stitched mesh over a regular grid to smooth for viewing.

IV. EXPERIMENTS

In order to validate the technique, a number of experiments were conducted.

A. AUV Platform

The University of Sydney’s Australian Centre for Field Robotics (ACFR), part of the ARC Centre of Excellence for Autonomous Systems, has a research class Autonomous Underwater Vehicle (AUV) called Sirius capable of undertaking high resolution survey work. This experimental platform is a modified version of a mid-size robotic vehicle called Seabed built at the Woods Hole Oceanographic Institution [18]. As shown in Figure 5, the submersible is equipped with a comprehensive suite of oceanographic sensors, including interferometric multibeam sonar, a CTD, Doppler Velocity Log (DVL) including a compass with integrated roll and pitch sensors, Ultra Short Baseline Acoustic Positioning System (USBL), forward looking obstacle avoidance sonar, and a high resolution stereo camera pair and strobes. The vehicle is controlled by an on-board computer which is used for sampling sensor information and running the vehicle’s low-level control algorithms. This platform is intended primarily as a research platform on which to test the novel navigation and sensing strategies being developed as part of this work. It is also ideally suited to low speed, high resolution imaging surveys and the data collected can be used to generate dense, high resolution terrain reconstructions of marine environments.

B. Mesh Experiments

We take a set of real multibeam sonar returns gathered using an AUV and generate a mesh based on the estimated
pose. The pose estimate is derived using a Kalman Filter [19] by fusing DVL, depth and GPS while the vehicle is on the surface and integrated DVL and depth while submerged. We first assess its performance by taking a known mesh and randomly rotating and translating it with respect to its origin. This allows us to verify that the technique is successfully able to recover the correct transformation given some arbitrary initial condition. Then the two independent passes are artificial rotated and translated by a random amount bounded by the estimated error in our ideal platform setup. We choose initial error values for x, y, z and roll, pitch, and yaw based upon what reasonable expected errors are between views based upon our sensors. We estimate the x, y error between separate passes to be quite large as the navigation pose sensing over long distances is inaccurate. This has been addressed previously with joining several local maps and then globally optimizing [20]. We believe that an inertial navigation based estimate of relative pose will be no better then 10m error in x and y between passes. Depth or z is known quite accurately due to a pressure sensor with an error of around 0.1m. So while x and y are quite unknown the relative z position is quite accurate for the purposes of registration. Finally relative roll and pitch are bounded around 3°, with yaw around 30° based upon the performance of the low cost tilt sensor with magnetic compass onboard the AUV. Using these as a rough guide we can attempt to register a mesh with itself under an artificial translation and rotation. Knowing the true correspondences we can estimate the error in our registration using the proposed technique.

One hundred iterations were performed with random starting offsets bounded by the values described above, we believe that is a sufficient number of runs to establish a consistent bound on the error performance. The standard deviation for each degree of freedom is shown in the following table. The values were calculated for the registration of a mesh to itself and then the registration of two separate passes of the same area. From this we can verify that the features adequately describe the mesh to the level where it can be registered almost exactly with itself under any artificial transform. Looking at Figure 4 we can see the fit between the two passes. This fit is quantified in Figure 7. Notice the constant offset that exists which we believe reflects navigation errors. The small standard deviations indicate the consistency of the algorithm. Looking at Figure 6(a) notice the improvement after the application of ICP over Figure 6(b). However no ground truth data exists for these two sets. The errors shown reflect not only errors in the registration but also the errors of the navigation solution and sampling resolution errors. Navigation based registration is an issue as these two meshes were taken from two distinct passes and from different depths. Additionally the effect of mesh resolution comes into play as the mesh is sampled at approximately one meter per cells, therefore features can only hope to be resolved and matched to around that level of granularity.

### V. CONCLUSIONS AND FUTURE WORK

We see this technique as integral to reconstructing environments in 3D. We believe this technique provides the framework necessary to register 3D data from any source. In future we hope to use this technique with heterogeneous data types. Through our experimentation we have learned that the uncertainties in the pose estimate are sufficiently bounded to register two meshes if there are enough salient features. We need to begin exploring if the features are independently descriptive, regardless of sensor modality, a requirement for achieving the registration of heterogeneous data. This will
registration

ICP.

Noticethat errors, the order of the algorithm application can result in errors.

However, utW/o w/oICP Table 7. Absolute value of 14C 18C 16C 14C 12C

18C

14C

12C

-0.30-0.24-0.12 0.41 0.51 0.85 0.16 0.43 0.49 0.63 0.16 0.13 2.58

Fig. 6. Absolute value of z Error in final registration with and without ICP. Notice there is a significant overall improvement in registration after the application of ICP. However this process leaves several small areas with large errors, which we attribute to the navigation errors and shadowing effects.

Fig. 7. Table containing the mean and standard deviations of the errors in registration of two separate passes in x,y,z in meters and \( \theta_x \), \( \theta_y \), and \( \theta_z \) in degrees. Notice there is a constant offset reflected in the means. However the algorithm settles on a fairly consistent registration with a standard deviation on the order of a cell size. We believe the constant offset reflects navigation errors and errors from the gridding process.

involve controlled experiments on data sets of the same region with different sensing modalities.

The exploration of how to insure the meshes are visually consistent is still underway. We are also looking at using visual features as a means of creating more visually consistent meshes as the visual textures are most apparent in the reconstructions. Finally, working on the global adjustment of errors in all registrations to maximize total mesh quality. Although there is a significant body of work dealing with the stitching and amalgamation of registered meshes [21], by improving the joining of the registered meshes the final aggregate mesh will more accurate. SLAM and structure from motion are great future directions that could produce excellent 3D map results improving upon this work’s local correspondences infavor of global optimizations.

We are working towards generating and matching meshes with multiple sensors. The data used can be gathered from any number of platforms operating at different heights from the seafloor, including an Autonomous Underwater Vehicle (AUV) and a ship or even Laser Airborne Depth Sounder (LADS) [22]. We see the future as combining visual imaging with sonar, which we believe will have the advantage of providing benthic context and overall depth profiles with the sonar, while high resolution stereo provides visual information for fine grain ecological assessment. The combination would provide a more complete reconstruction than either photo mosaics or multibeam bathometry alone. A multibeam sonar and pair of stereo vision cameras provide a strong contrast to one another in operating range and resolution. Acoustic imaging in underwater environments, specifically multibeam sonar, provides significantly longer range data at lower resolution than optical imaging with a standard lens. But both technologies in tandem is an excellent method of utilizing sensor fusion to improve benthic survey.

ACKNOWLEDGMENT

We would like to thank the captain and crew of the Melantre and the Australian Center for Field Robotics technical staff, especially Ritesh Lal, Jeremy Randle and Alan Trinder. This work is supported by the ARC Centre of Excellence Programme, funded by the Australian Research Council (ARC) and the New South Wales State Government.

REFERENCES


