Underwater environment reconstruction using stereo and inertial data

Andrew Hogue
Faculty of Business and Information Technology
University of Ontario Institute of Technology
Oshawa, Ontario, Canada
andrew.hogue@uoit.ca

Andrew German and Michael Jenkin
Department of Computer Science and Engineering
The Centre for Vision Research
York University, Toronto, Ontario, Canada
{german.jenkin}@cse.yorku.ca

Abstract—The underwater environment presents many challenges for robotic sensing including highly variable lighting, the presence of dynamic objects, and the six degree of freedom (6DOF) 3D environment. Yet in spite of these challenges the aquatic environment presents many real and practical applications for robotic sensors. A common requirement of many of these tasks is the need to construct accurate 3D representations of structures in the environment. In order to address this requirement we have developed a stereo vision-inertial sensing device that we have successfully deployed to reconstruct complex 3D structures in both the aquatic and terrestrial domains. The sensor temporally combines 3D information, obtained using stereo vision algorithms with a 3DOF inertial sensor. The resulting point cloud model is then converted to a volumetric representation and a textured polygonal mesh is extracted for later processing. Recently obtained underwater reconstructions of wrecks and coral obtained with the sensor are presented.

I. INTRODUCTION

The underwater environment presents numerous challenges for the design of robot vision sensors, and algorithms. Yet it is these constraints and challenges that make this environment almost ideal for the development and evaluation of robotic sensing technologies. Vehicles and their sensors operating in this environment must cope with unknown motion due to external forces such as currents, surf, and swells. These actions can produce unwanted and unpredictable 6DOF motion of the sensor. In spite of these complexities, the underwater environment provides a vast range of applications for which environmental reconstruction is desirable including inspection and entertainment applications. The technology is also applicable to a wide variety of terrestrial applications from crime scene analysis to rapid 3D prototyping for movie special effects.

A. Underwater Inspection

Underwater inspection tasks find application in areas as diverse as ensuring security of hull vessels in ports, to monitoring the health of coral reefs. A robot capable of autonomously reconstructing a 3D model of a ship’s hull would enable inspection and analysis of its structural integrity and the search for malicious payloads. Similarly, a 3D model of a ship wreck could aid in determining the cause of the event. In addition to man-made structures, there is significant need for the analysis of natural structures, such as coral reefs. Coral reefs are important for the global underwater ecosystem. They provide both shelter and nutrition for many species of aquatic life. Reefs are composed of extremely fragile organisms that can be destroyed by slight changes in water temperature, salinity and pollution. One metric for establishing the health of the reef is to monitor its size with respect to time. Manually measuring reef size is an error prone and time consuming effort. Another indicator of reef health is the variety of coral and surrounding aquatic life. Automatic 3D reconstructions can be used to analyse the health of the reef by measuring changes to both the size of the reef, and the population density of the coral itself.

B. Aquatic Entertainment

The aquatic environment provides a range of recreational activities. Potential applications in the entertainment industry of 3D reconstruction technology includes the reconstruction of underwater scenes for documenting recreational scuba diving expeditions. Many divers videotape their dives using traditional consumer-grade video recording equipment. Automatic 3D reconstruction of the dive could add impact to the experience by increasing the dimensionality of the recording without increasing task-loading on the diver.

C. Land Applications

Although the work described here is directed primarily at the aquatic domain, many land-based reconstruction applications exist. For example, the 3D analysis of crime scenes can provide spatially significant information that is unavailable in traditional photographic evidence. This information is useful for both investigative purposes and also for later courtroom demonstrations. Another application can be found in accessibility assessment. The evaluation of home accessibility for the disabled is a time-consuming but necessary task. 3D scene reconstruction could be used to provide more effective and accurate assessment. Finally, in the entertainment industry, 3D scene reconstruction enables the production of low-cost 3D animations which can be quickly created for the pre-visualization stage of movie making.

II. ENVIRONMENTAL RECONSTRUCTION

A natural sensing choice for autonomous aquatic scene reconstruction is to use cameras and computer vision techniques. In the terrestrial domain, sensing technologies such as stereo vision coupled with good vehicle odometry has been used to construct 2D top-down maps and 3D models of the environment. There is a long history of research in simultaneous localization and mapping (SLAM) for robotic
vehicles (e.g. [1], [2], [3], [4]). Terrestrial SLAM algorithms often assume a predictable vehicle odometry model to assist in the probabilistic association of sensor information with salient visual features. The lack of such predictable vehicle odometry in the underwater domain necessitates solutions which are more dependent upon sensor information than is traditional in the terrestrial domain. This has prompted recent research in robotic vehicle design, sensing, localization and mapping for underwater vehicles (see [5], [6], [7], [8]).

A vehicle that requires robust and versatile sensing strategies to extract 3D models of the environment is the AQUA robot [9]. The AQUA robot is a visually guided autonomous robot developed through a collaboration between researchers at Dalhousie University, McGill University and York University. The vehicle is a versatile hexapod capable of amphibious operation. On land, its legs provide foot-ground contact that propel the vehicle. Underwater these same legs act as flippers or fins to drive the vehicle both along the surface of the water and at depths to 30 metres. In order to meet the sensing needs of the AQUA robot, a series of visual-inertial sensors have been built[10]. The primary goal of these sensors is to collect high framerate multi-camera video imagery coupled with synchronized time-stamped inertial information. This data is later processed off-line to obtain 6DOF ego-motion and 3D models of the environment. Although we describe the standalone sensor package here, the long-term goal is to incorporate the sensing hardware and algorithms into the AQUA vehicle itself.

III. THE AQUASENSOR

The hardware design goal for the AQUASENSOR was to construct a compact, fully-contained unit that is sufficiently portable to be used by a single diver and whose components, suitably repackaged, could later be integrated within the AQUA robot. We combine range information extracted from stereo imagery with 3DOF orientation from an inertial measurement unit (IMU) within a SLAM algorithm to accurately estimate a dense 3D model of the environment and the trajectory of the sensor. It is important to note that although the AQUASENSOR does not process the collected data in realtime, the algorithms used to analyze the data are being developed with realtime performance in mind.

The AQUASENSOR is sealed within a custom underwater housing that permits operation of the device to a depth of 30 metres. The sensor is operated either via an IEEE 802.11g wireless link for surface operation, or via waterproof switches mounted on the exterior of the sensor when the device is deployed underwater. The wireless interface onboard the sensor provides status information and sample imagery to offboard devices. LEDs mounted on the exterior of the sensor provide status information to the operator. The LEDs and waterproof switches are the only communication possible with the device when submerged.

IV. DATA PROCESSING

Upon return to the surface, data from the AQUASENSOR is offloaded to higher performance computers and a larger disk array for processing. Recovering accurate depth from
stereo imagery is performed by leveraging the optimized sum-of-squared differences algorithm implemented in the Point Grey Triclops library\(^1\). This provides a dense set of 3D points per acquired image frame. To estimate the motion of the camera we register point sets from different times into a common reference frame. Using the 2D image motion, we estimate the change in the 6DOF position and orientation of the sensor. This is accomplished by tracking interesting features temporally, using only features that correspond to estimated 3D locations, and computing the position and orientation change between the sequential frames using a least-squares algorithm. We integrate the change in pose each frame which means that any error in the estimate will accumulate over time. Also, since the intrinsic calibration parameters for the cameras are not known perfectly, any error caused by calibration error will also accumulate causing the trajectory to diverge from reality. To reduce this error, we utilize a 3DOF inertial measurement unit to provide complimentary information about the orientation of the device.

A. Visual Ego-motion estimation

First, “good” features are extracted from the reference camera at time \(t\) using the Kanade-Lucas-Tomasi feature tracking algorithm (see [12], [13]) and are tracked into the subsequent image at time \(t + 1\). Using the disparity map previously extracted for both time steps, tracked points that do not have a corresponding disparity at both time \(t\) and \(t + 1\) are eliminated. Surviving points are subsequently triangulated to determine the metric 3D points associated with each disparity.

In underwater scenes, many objects and points are visually similar and thus many of the feature tracks will be incorrect. Dynamic illumination effects, aquatic snow, and moving objects (e.g. fish) increase the number of spurious points that may be tracked from frame-to-frame. To overcome these problems, we employ robust statistical estimation techniques to label the feature tracks as belonging to either a static or non-static world model. This is achieved by estimating a rotation and translation model under the assumption that the scene is stationary. The resulting 3D temporal correspondences are associated with stable scene points for the basis of later processing.

We represent the camera orientation using a quaternion and compute the least-squares best-fit rotation and translation for the sequence in a two stage process. First, using RANSAC [14] we compute the best linear least squares transformation using Horn’s absolute orientation method [15]. Given two 3D point clouds \(r_{t_0}\) and \(r_{t_1}\) at times \(t_0\) and \(t_1\) respectively, we estimate the rotation and translation to bring \(r_{t_1}\) into accordance with \(r_{t_0}\). The centroid, \(\bar{r}_{t_0}\), of each point cloud is computed and subtracted from the points to obtain two new point sets, \(r'_{t_0} = r_{t_0} - \bar{r}_{t_0}\) and \(r'_{t_1} = r_{t_1} - \bar{r}_{t_1}\).

To compute the rotation, \(R(\cdot)\), we minimize the error function

\[
\sum_{i=1}^{n} \left| \left| r'_{t_0,i} - sR(r'_{t_1,i}) \right| \right|^2
\]

The rotation, \(R(\cdot)\), and scale, \(s\), are estimated using a linear least-squares approach (detailed in [15]). After estimating the rotation, the translation is estimated by transforming the centroids into a common frame and subtracting.

Noting that the above method is prone to an imperfect result, the final step is to refine the rotation and translation simultaneously using a nonlinear Levenberg-Marquardt minimization [16] over six parameters. For this stage we parameterize the rotation as a Rodrigues vector [17] and estimate the rotation and translation parameters by minimizing the transformation error

\[
\sum_{i=1}^{n} \left| \left| r_{t_0,i} - (R(r_{t_1,i}) + T) \right| \right|^2
\]

In practice, we find that the minimization takes few iterations to minimize the error to acceptable levels which does not necessarily preclude realtime operation. This approach to pose estimation differs from the traditional Bundle-Adjustment approach [11] in the structure-from-motion literature in that it does not refine the 3D locations of the features as well as the trajectory. We chose not to refine the 3D structure to limit the number of unknowns in our minimization and thus provide a solution to our system more quickly.

B. IMU Integration

Egomotion estimation via vision motion introduces at least two sources of error in the estimate of 6DOF pose. First, the point-cloud registration computed from feature tracks can never be perfect. As such, there is always a small residual error in the registration per frame which accumulates over time. Second, the intrinsic camera parameters are not known perfectly and any error in these parameters introduces an error in each 3D point. In particular, the radial distortion estimate of the lens is prone to error and the per-point error is non-uniform over the visual field. This can introduce an artificial surface curvature in the 3D point clouds which is subtle but noticeable when many frames are registered. This effect can be seen in Figure 2(a). Here, the registration error is small and the points line up very well creating a visually “correct” model when viewed closely, however after registering many frames it can be seen that there is an introduced curvature to the recovered surface. To help counteract this effect, we utilize the 3DOF IMU to provide more information about the orientation of the device. The IMU provides us with a quaternion representing the absolute orientation in 3D space. We enforce the fact that the change in orientation as computed from the visual system must be consistent with the change in the absolute orientation of the device. This is accomplished by transforming the IMU orientation change

\[
\text{IMU} \ q_6 = \text{IMU} \ q_{t-1} \ast \text{IMU} \ q_t
\]

\(^1\)http://www.ptgrey.com
into the estimated camera frame,
\[ q_t = \frac{C_{AM} q_0}{C_{AM}} q_{t-1}^{\text{imu}} q_8 \]  
(4)
and then performing the Levenberg-Marquardt minimization using this new pose as the initial guess. The effect of utilizing this information can be seen in Figure 2(b) where the curvature has been reduced in the resulting model.

C. Reconstruction Algorithm

The reconstruction algorithm is summarized by the following:

1) Perform the stereo disparity extraction algorithm and estimate 3D point cloud at time t
2) Track salient features from time t to t – 1
3) Prune 2D features to only the ones with a corresponding 3D point
4) Estimate 3D vision-only pose change using RANSAC, \( C_{AM} q_8, T_t \)
5) Compute \( q_8^{\text{imu}} \) and \( q_t \) as above
6) Refine \( q_t, T_t \) using Levenberg-Marquardt minimization for only a few iterations
7) Apply pose to point cloud and add to octree data structure
8) Extract mesh using a constrained elastic surface-net algorithm [18] or the Marching Cubes algorithm[19].

V. EXPERIMENTS

Experiments have been performed to both evaluate the accuracy of the reconstruction system and to create 3D models of real-world objects in the field. Results from field experiments near Holetown, Barbados show the reconstruction of a coral reef growing on the surface. Figure 3(b) shows a polygon mesh representation of the model obtained with the Marching Cubes Algorithm. Figure 3(c) shows a small section of a sunken barge lying in Folkstone Marine Reserve. Sample qualitative reconstructions from the underwater sequences are shown in Figure 3. Figure 3(a) shows a small section of a model of a sunken barge with coral reef growing on the surface. Figure 3(d) shows a 2D mosaic created by imaging the barge manually and swimming along the same trajectory the sensor followed.

Land-based reconstructions demonstrate the ability to use the sensor to reconstruct terrestrial scenes captured with 6DOF hand-held motion. To evaluate the accuracy of the sensor, we performed experiments and compared the recovered trajectory against ground truth trajectory data. To acquire ground truth 3D data, we constructed a scene with known measurements and independently tracked the camera sensor with an IS-900TM motion tracking system from Intersense® (Figure 4). The average RMS error of the sensor trajectory over a distance of 3 metres is approximately 2cm and an error of 0.5% for the reconstructed close range 3D data.

VI. DISCUSSION AND FUTURE WORK

Traditional underwater sensing devices have relied on active sensors (sonar in particular) to recover three-dimensional environmental structure. Advances in stereo sensing and data fusion technologies demonstrates that passive stereo is a sufficiently robust technology to be applied in the aquatic domain as well. Although the underwater domain presents unique challenges to traditional vision algorithms, robust noise-suppression mechanisms can be used to overcome many of these difficulties. Results from land-based experiments demonstrate the accuracy of our reconstruction system, and more experiments to assess the accuracy underwater are ongoing. Results from underwater field trials also functionally demonstrate the system’s ability to reconstruct qualitatively correct 3D models of the aquatic environment. The entire reconstruction system operates at 2-4 frames per second (depending on the stereo algorithm parameter settings). The current speed limitation is due to the stereo
Fig. 3. Underwater reconstructions of a sunken barge with coral growth. (a) shows the recovery of a small region of the barge. Upper left panel shows the current surface model as a point cloud. Upper right panel shows the current 3D matched points that are used to estimate ego motion. Lower two panels are the raw stereo inputs. (b) shows the recovered polygonal mesh along with the camera egomotion. (c) shows the model of a large section (18 metres) of the barge as a point cloud with inset images to show close-up views of the 3D model that illustrate scale and detail. A 2D mosaic of the barge (created manually) is shown in (d) with the recovered section enclosed in a red box to place the 3D model in context.
Algorithm and feature tracking which could be implemented in graphics hardware at near real-time rates. Recovery of a long (18 metre) section of sunken barge is shown containing approximately 35 million 3D points. To place the model in context, we have shown a manually registered 2D mosaic from images of the same barge and highlighted the area we recovered in 3D. The sensing system described in this paper is part of the AQUA robot sensor package and can also be used as a standalone single-diver-deployable sensor for coral reef analysis. Future work includes embedding the technology within the robot housing and developing a more robust SLAM algorithm that incorporates large loop-closing to further reduce the pose error over long distances.

ACKNOWLEDGMENTS

We gratefully acknowledge the funding provided by NSERC and IRIS NCE for the AQUA project. We also would like to thank Sunbir Gill for help with the meshing software, Jim Zacher and the McGill AQUA team for engineering support and the McGill Bellairs Research Institute for providing a positive atmosphere during the field research trials. We also thank Urszula Hogue and Jennifer Lamb for their patience and proofreading.

REFERENCES