Development of an Underwater Vision Sensor for 3D Reef Mapping

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

Abstract—Coral reef health is an indicator of global climate change and coral reefs themselves are important for sheltering fish and other aquatic life. Monitoring reefs is a time-consuming and potentially dangerous task and as a consequence autonomous robotic mapping and surveillance is desired. This paper describes an underwater vision-based sensor to aid in this task. Underwater environments present many challenges for vision-based sensors and robotic vehicles. Lighting is highly variable, optical snow/particulate matter can confound traditional noise models, the environment lacks visual structure, and limited communication between autonomous agents including divers and surface support exacerbates the potentially dangerous environment. We describe experiments with our multi-camera stereo reconstruction algorithm geared towards coral reef monitoring. The sensor is used to estimate volumetric scene structure while simultaneously estimating sensor ego-motion. Preliminary field trials indicate the utility of the sensor for 3D reef monitoring and results of land-based evaluation of the sensor are shown to evaluate the accuracy of the system.

I. INTRODUCTION

Small changes in climate can produce devastating results on the world’s coral reef population\(^1\). For example, an increase in ocean temperatures of only a few degrees destroyed most of the coral in Okinawa in 1998\(^1\). Thus, coral reefs are an excellent indicator of global climate change. Monitoring reefs can be a very labor-intensive task. An international organization called Reef Check\(^2\) was established in 1997 to frequently and accurately monitor coral reef environments. The methods used by Reef Check rely on hundreds of volunteer divers to identify aquatic species over 100m transects. This task is very time-consuming, error-prone, and is potentially dangerous for the divers.

Advances in computer vision and robotic technology can be used to aid divers in monitoring tasks performed by organizations such as Reef Check. The development of an autonomous robot to survey a particular reef area or transect would be of enormous value. After a reef section has been selected for monitoring, the robot could be placed near the reef and could then travel autonomously along the designated transect collecting data for later analysis. Algorithms such as \([2]\) or \([3]\) could then be used for the automatic classification of coral reefs. The raw data stream could be stored for later viewing by a human operator to identify and categorize aquatic life.

Alternatively, automatic algorithms could be developed for fish classification in a similar vein to the coral reef classifiers noted above. After the data has been analyzed, the robot could be re-deployed over the same site to enable long-term monitoring of a particular coral reef. An autonomous tool such as this would facilitate a Reef Check paradigm that is much safer to deploy and would provide a bias-free and less dangerous long-term monitoring solution.

To this end, this paper describes an underwater sensor capable of generating 3D models of underwater coral reefs. Our sensor is designed to be deployed and operated by a single diver and ongoing research is investigating full integration of the sensor with AQUA\(^4\), an amphibious underwater robot (see Figure 1). The sensor is capable of collecting high-resolution video data, generating accurate three-dimensional models of the environment, and estimating the trajectory of the sensor as it travels. The algorithm developed is primarily used offline to extract highly accurate 3D models, however if low-resolution images are used, the system operates at 7fps using off-the-shelf hardware.

The underwater environment presents numerous challenges for the design of robotic systems and vision-based algorithms. Yet it is these constraints and challenges that make this environment almost ideal for the development and evaluation of robotic and sensing technologies. Vehicles operating in this environment must cope with the potentially unpredictable six-degree-of-freedom (6DOF) motion of the vehicle. Currents, surf and swell can produce unwanted and unexpected vehicle motion.

A natural choice for sensing for an aquatic vehicle is to use cameras and computer vision techniques to aid in navigation and trajectory reconstruction. Unfortunately a host of problems plague underwater computer vision techniques. For example, the turbidity of the water caused by floating sedimentation (“aquatic snow”) and other floating debris can cause problems for standard techniques for visual understanding of the world. The optical transmittance of water varies with wavelength and dynamic lighting conditions can make it difficult to reliably track visual features over time. Dynamic objects such as fish and coral can create spurious measurements influencing pose-estimation between frames. These challenges have prompted recent research in robotic vehicle design, sensing, localization and mapping for underwater vehicles (cf. \([5]\), \([6]\), \([7]\), \([8]\)).

\(^1\)http://www.marinobiology.org/coralbleaching.htm
\(^2\)http://www.reefcheck.org
In the terrestrial domain, sensing technologies such as stereo vision coupled with good vehicle odometry has been used to construct 3D environmental models. Obtaining such odometry information can be difficult without merging sensor data over the vehicle’s trajectory, and as a result there is a long history of research in simultaneous localization and mapping (SLAM) (cf. [9], [10], [11], [12]) for robotic vehicles. Terrestrial SLAM algorithms often assume a predictable vehicle odometry model to assist in the probabilistic association of sensor information with environment features. The lack of such predictable vehicle odometry underwater necessitates solutions which are more dependent upon sensor information than is traditional in the terrestrial domain. The algorithm presented here utilizes passive stereo vision to obtain local depth estimates and uses temporal feature tracking information to estimate the vehicle trajectory. Experimental results obtained with the algorithm during recent sea trials illustrate the effectiveness of the approach but are difficult to evaluate objectively. Land-based reconstructions conducted using the same algorithm and hardware are presented to evaluate the accuracy of the system against ground-truth trajectory data.

II. THE AQUA ROBOT

The algorithm described here is designed to operate under manual operation and also as a component of the AQUA robot (see [4], [13], [14], [15] and Figure 1). AQUA is a six-legged amphibious robot based on a terrestrial hexapod named RHex [16]. RHex was the product of a research collaboration between the Ambulatory Robotics Lab at McGill, the University of Michigan, the University of California at Berkeley and Carnegie Mellon University. AQUA differs from RHex in its ability to swim. AQUA has been designed specifically for amphibious operation and is capable of walking on land, swimming in the open water, station keeping at depths to 15m and crawling along the bottom of the ocean. The vehicle deviates from traditional ROV’s in that it uses legs for locomotion. When swimming, the legs act as paddles to displace the water and on land the legs allow AQUA to walk on grass, gravel, sand, and snow. The paddle configuration gives the robot direct control over five of the six degrees of freedom: surge, heave, pitch, roll and yaw. An inclinometer, forward and rear facing video cameras, and an on-board compass are used to assist in the control of the robot’s motion.

A. The AQUASENSOR

We have developed three versions of a standalone sensor package (see Figures 2 and 3) that is to be integrated into the AQUA vehicle in the future. The configuration and limitations of each version of the sensor are discussed briefly here and the reader is referred to [17] for further details. AQUASENSOR V1.0 consists of three fully calibrated Firewire (IEEE1394a) cameras in an L-shaped trinocular configuration used to extract dense depth information from the underwater environment. It also integrates a 6DOF inertial measurement unit (Crossbow IMU-300CC) to maintain relative orientation and position of the unit as it moves, and the necessary electronics to drive the system. The cameras and inertial unit are tethered to a base computer system that is operated on land or on a boat via a custom-built 62m fibre-optic cable. The cameras capture 640x480 resolution gray-scale images at 30 frames per second and all three video streams are saved for later analysis. Prior to operation, the intrinsic and extrinsic camera parameters are estimated using Zhang’s camera calibration algorithm [18] and an underwater calibration target. AQUASENSOR V1.0 was field tested in Holetown, Barbados and hundreds of gigabytes of trinocular and synchronized IMU data have been captured using the device.

Cable management and mobility were issues in the field with this version of the sensor. This initial version of the sensor required significant surface support personnel to manage the long tether, two divers were needed to orient and move the sensor housing, and weather conditions were problematic for the base unit and operator.

Building on our previous design with new mobility constraints in mind, we re-designed the entire sensor to accommodate a more mobile solution that a single diver could deploy and operate.

For AQUASENSOR V2.0 and V2.1 we adopted a Point Grey Bumblebee™ to capture colour stereo images which had the added benefit of a smaller footprint than our previous camera system. Due to the smaller size of the Bumblebee and the adoption of a smaller inertial unit (a 3DOF Inertiacube3™ from Intersense4) we were able to re-design the

---

3http://www.ptgrey.com

4http://www.isense.com
camera system to be smaller and simpler to operate. We placed all computing hardware into our previous underwater housing and connected the cameras to the computing unit via a short (1.5m) cable. This new sensor design eliminated issues with tether management and removed the need for surface support personnel. The computing unit could be harnessed directly to the diver and the neutrally buoyant hand-held unit was much simpler to orient towards reef and other structures for data collection by a single diver.

### III. Mapping Algorithm

For AQUASENSOR V1.0, we utilized a hierarchical approach for stereo disparity extraction based upon the real-time algorithm of Mulligan et al. [19]. After rectification, an image pyramid is created and the stereo matching algorithm is performed over the specified disparity range beginning at the coarsest level of the pyramid. An occlusion map is extracted by performing the stereo algorithm in the reverse direction on the left-right image pair for the coarsest level. For each level in the pyramid, the disparity map is transformed to the appropriate resolution and the disparities are refined using the successive pyramid levels. Example disparity images are shown in Figure 2. AQUASENSOR V2.0 leveraged the real-time sum-of-squared differences algorithm implemented in the Point Grey Triclops library to extract disparity maps more efficiently.

#### A. Ego-motion estimation

In order to estimate the motion of the camera, we utilized both 2D image motion and 3D data from the extracted disparities. First, “good” features are extracted from the left camera at time $t$ using the Kanade-Lucas-Tomasi feature tracking algorithm (see [20], [21]) 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. The 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 and moving objects (e.g. fish) increase the number of incorrect points that are tracked from frame-to-frame. To overcome these problems, we employ robust statistical estimation techniques to label the feature tracks as either static or non-static. 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[22] we compute the best linear least squares transformation using Horn’s absolute orientation method[23]. 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, $\bar{r}_{t_0}, \bar{r}_{t_1}$, of each point cloud are 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, $R(\cdot)$ represented as a quaternion, we minimize the error function

$$\sum_{i=1}^{n} \| r'_{t_0,i} - sR(r'_{t_1,i}) \|^2.$$  

The rotation, $R(\cdot)$, and scale, $s$, are estimated using a linear least-squares approach (detailed in [23]). After estimating the rotation, the translation is estimated by transforming the centroids into a common frame and subtracting.

The final step is to refine the rotation and translation using a nonlinear Levenberg-Marquardt minimization[24] over six parameters. For this stage we parameterize the rotation as a Rodrigues vector[25] and estimate the rotation and translation parameters by minimizing the transformation error

$$\sum_{i=1}^{n} \| r_{t_0,i} - (R(r_{t_1,i}) + T) \|^2.$$  

In practice, we found that the minimization takes less than five iterations to minimize the error to acceptable levels. This approach to pose estimation differs from the traditional Bundle-Adjustment[26] in the structure-from-motion literature in that it does not refine the 3D locations of the features. 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. Our minimization is thus prone to

---

Fig. 3. Evolution of the AQUASENSOR design. V2.0 (a-c) and V2.1(d-e). (a) Hand-held unit (b) Hand-held unit with Bumblebee (c) Complete unit (d) Final system (e) Revised Hand-held unit.
error in the 3D structure which is correlated with the error in the pose. We overcome this limitation by utilizing Point Grey’s error model of the disparity estimation on a per point basis. The accuracy of each 3D point is determined from the stereo parameters used in the disparity estimation process this information and is used in the 3D modeling phase. We plan on utilizing this information as part of a weighted least-squares estimation of pose in the future.

B. Volumetric Reconstruction

We use a volumetric approach to visualize the resulting 3D model. We register the 3D point clouds into a global frame using the previously computed camera pose and each point is added to an octree. Upon adding the points to the octree, they are averaged to maintain a constant number of points per node. The octree is pruned to remove isolated points which produces a result that is less noisy in appearance and can be manipulated in real-time for visualization. The octree can be viewed at any level to produce a coarse or fine representation of the underwater data. Subsequently, a mesh can then be extracted using algorithms such as the constrained elastic surface net algorithm[27].

IV. EXPERIMENTS

In this section, we describe the evaluation of our reconstruction system. We performed several experiments and field trials to demonstrate the utility and accuracy of the sensor. Results from field experiments show the reconstruction of a coral bed and pieces of a sunken barge found in Folkstone Marine Reserve in Holetown, Barbados. Sample images from the underwater sequences are shown in Figure 4 and their corresponding reconstructions can be seen in Figure 5. The recovered camera trajectory is shown in the reconstructions as well. In figures 5(a) and 5(b), the structure of the rope and hole in the barge is clearly visible. The second sequence (figures 5(c) and 5(d)) is of a bed of coral growing on the barge.

Land-based reconstructions show 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 to 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 Intersense5. We also manually measured the distance between already identified points in the physical scene using a standard measuring tape and augmented our system to allow for interactive picking of 3D points in the model.

We placed two yellow markers 63.5 cm apart on a platform and moved the sensor in an approximate straight line from one marker to the other. Figures 6(a) and 6(b) show the stereo pairs associated with the first and the last image of the sequence. A second test was also constructed in which the sensor was moved in a circular trajectory (Figure 6(f)-(h)). An Intersense IS-900 tracking system was used to provide absolute pose and the pose error was computed on a frame to frame basis.

We performed a manual calibration of the offset between the IS-900 sensor and the camera coordinate frame to bring

5http://www.isense.com
the measurements into the same space. The euclidean distance between the 3D camera positions of the vision-based reconstruction and the IS-900 absolute position was used as an error metric and the resulting error is plotted in Figure 7 for the two reconstructions. In the first error plot, the mean error was found to be 2.1 cm and in the second the mean error was 1.7 cm. The reconstructed 3D models for these experiments are shown in Figure 6. The error is computed as the euclidean distance between the position estimate given from our system versus the absolute position given from the IS-900 tracker.

V. 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 now 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. Results from underwater field trials also demonstrate that the system can operate underwater to reconstruct structures underwater.

The sensing system described in this paper is part of the AQUA robot sensor package. Future work includes the embedding of the technology within the robot housing and integrating the stereo reconstruction with inertial data to improve frame-to-frame motion estimation when no texture is available.

ACKNOWLEDGMENTS

We graciously acknowledge the funding provided by NSERC and IRIS NCE for the AQUA project. We also would like to thank 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 Jasiobedzka for her patience and proofreading.

REFERENCES