AQUA: an aquatic walking robot

Christina Georgiades*, Andrew German**, Andrew Hogue**, Hui Liu‡, Chris Prahacs*,
Arlene Ripsman**, Robert Sim‡‡, Luz-Abril Torres‡‡, Pifu Zhang‡
Martin Buehler*, Gregory Dudek‡‡, Michael Jenkin**, Evangelos Milios‡

* Mechanical Engineering
McGill University
3480 University St.
Montreal, PQ, Canada

**Computer Science and Engineering
York University
4700 Keele St.
Toronto, Ontario, Canada

‡Faculty of Computer Science
Dalhousie University
6050 University Ave.
Dalhousie, NS, Canada

‡‡School of Computer Science
McGill University
3480 University St.
Montreal, PQ, Canada

{cgeorg, simra, latorres, cprahacs, buehler, dudek}@cim.mcgill.ca,
{aagerman, hogue, arlene, jenkin}@cs.yorku.ca,
{hliu, pifu, eem}@cs.dal.ca

***Abstract* – This paper describes an underwater walking robotic system being developed under the name AQUA, the goals of the AQUA project, the overall hardware and software design, the basic hardware and sensor packages that have been developed, and some initial experiments. The robot is based on the RHex hexapod robot and uses a suite of sensing technologies, primarily based on computer vision and INS, to allow it to navigate and map clear shallow-water environments. The sensor-based navigation and mapping algorithms are based on the use of both artificial floating visual and acoustic landmarks as well as on naturally occurring underwater landmarks and trinocular stereo.**

Keywords-autonomous robot, aquatic robot, robotic sensing

I. INTRODUCTION

Mobile robotics is frequently cited as being most appropriate for application domains that are costly, inconvenient, or inhospitable for humans to work in. The aquatic domain is an almost perfect fit. The environment is dangerous, and many tasks require long-term operation and significant depth. Mobile robotics is particularly well suited to underwater applications such as reef or pipeline inspection, fish stock surveillance, marine life observation and environmental disaster assessment.

Many of these applications involve stationary observation. That is, although mobility is required to get the vehicle to the close proximity of the task, the task itself relies on the vehicle to maintain a constant pose (often near or on solid objects in the environment). Unlike the terrestrial domain, in which station keeping may be as simple as powering down the locomotion system, in the aquatic domain station keeping is a complex task. A thruster-driven aquatic robot must actively and continually control its thrusters and buoyancy in order to maintain its pose. In addition to the obvious energy consumption issue associated with this active station keeping, thrusters operated near the sea bottom may disturb sand and other debris, reducing the ability of sensors.

A second issue with thruster-based aquatic vehicles is that these vehicles can only operate in the water. That is, they must be deployed and recovered from sufficiently deep water for the vehicle to be able to maneuver. Surveying/inspection in shallow water, or deploying the vehicle from the beach is not possible for traditional aquatic vehicles.

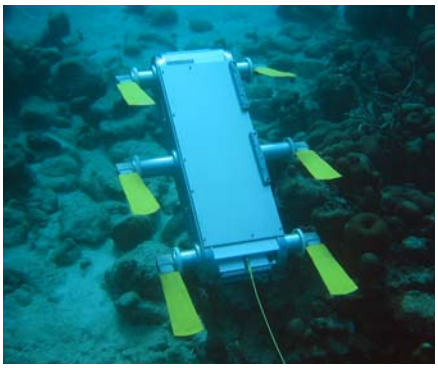
In order to address these issues, we are developing an aquatic walking robot -- AQUA. Through an appropriate design of limbs for the vehicle, the vehicle's legs can be used for both traditional walking locomotion strategies (either on land, or along the bottom of the aquatic environment), as well as to propel the vehicle through the water by swimming.

Developing a walking aquatic robot requires the solutions to fundamental issues related to locomotion, sensing, navigation and reasoning. Many existing approaches to these classic robotic tasks are not directly applicable to the aquatic walking environment and some applicable techniques entail important new challenges in the aquatic domain.

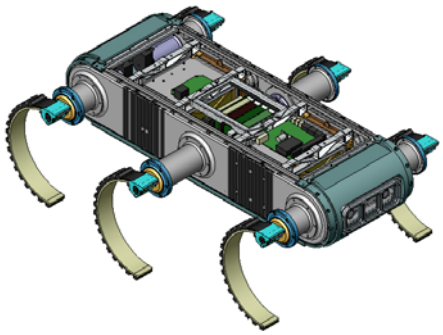
This paper describes the basic approaches that are being taken within the AQUA project to develop a fully autonomous walking aquatic vehicle. It describes the basic design of the locomotive and sensing hardware, and describes initial results in terms of vehicle locomotion and sensing.

II. THE SASR TASK

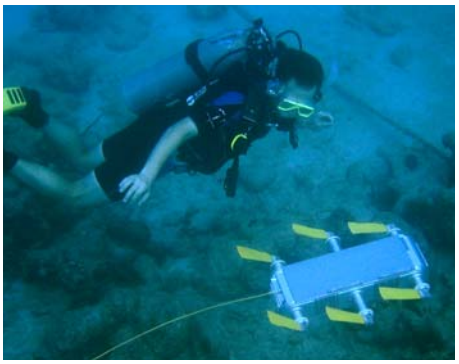
This Site Acquisition and Scene Re-inspection (SASR) task is fundamental to many underwater robotic tasks. A typical scenario in a SASR task is as follows. A robot is deployed near the site, in our case on a nearby beach. Under operator control or supervision, the robot walks out into the water and is controlled or directed to a particular location on the seabed where sensor measurements are to be made. (This may be the supposed location of some environmental incident, the



(a) The robot at sea



(b) Sketch of the robot with legs



(c) Robot with diver

Figure 1. The AQUA robot. (a) shows the robot swimming over a coral reef. The vehicle has six fins (or legs), which can be independently controlled. Here the vehicle is swimming up towards the surface while being tethered to an external operator. (b) shows the arrangement of internal components, and treaded legs for use in walking on shore and/or on the bottom. (c) shows the robot with a diver for scale.

location of known fish stocks that must be inspected periodically, or some similar task.) Once near the required site, the robot navigates to the selected location, where it utilizes its ability to move as a walking vehicle to achieve an appropriate pose from which to undertake extensive sensor readings over an extended time period. Once measurements are made, the robot then returns home autonomously. Later, the robot actively guides – and potentially controls – its motion to the previously visited site in order to collect additional data. One key target application we are examining relates to the regular inspection and monitoring of fragile marine ecosystems where unobtrusive observation over substantial time periods is appropriate.

Solving the SASR task requires solutions to a number of scientific and engineering questions including problems of position and pose estimation in unstructured environments, underwater landmark recognition, robotic navigation, motion control, path planning, vehicle design, environment modeling and scene reconstruction, 3D environment exploration and autonomous and teleoperated control of an aquatic legged vehicle. Here we describe some of the results to date in the search for solutions to these problems.

III. THE HARDWARE

A. *The Vehicle*

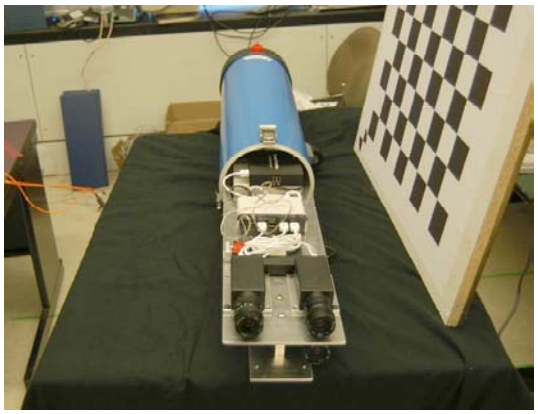
AQUA is an aquatic robot capable of both legged and swimming motion (see Figure 1(a)). AQUA is based on RHex, a terrestrial six-legged robot developed in part by the Ambulatory Robotics Lab at McGill in collaboration with the University of Michigan, the University of California at Berkeley and Carnegie Mellon University [1] (see figure 1(b)). AQUA's required capabilities are surface and underwater swimming, diving to a depth of 10m, station keeping and crawling at the bottom of the sea. For propulsion, the vehicle does not use thrusters, as do most underwater vehicles. Instead it uses six paddles, which also act as control surfaces during swimming, and as legs when walking. The paddle configuration gives the robot direct control over five of the six degrees of freedom that it has: surge (back and forth), heave (up and down), pitch, roll and yaw. An inclinometer and a compass onboard are used in the control of the robot's motion underwater.

The robot is approximately 65 cm long, 50 cm wide (at the fins), and 13cm high. It has an aluminum waterproof shell and displaces about 18 kg of water. A buoyancy control system is currently being developed, making the robot negatively, neutrally or positively buoyant. The robot is also equipped with a rail on which can be mounted the trinocular sensor package or the acoustic localization system (described below). The robot is power autonomous. Two onboard NiMH batteries provide over two hours of continuous operation. Signals from cameras mounted within the AQUA vehicle itself, from the sensor systems mounted on the robot, as well as the command and control output, are brought to a floating platform at the surface via a fiber optic tether. A wireless link exists between the platform and a shore-based operator. The operator uses the information from the onboard cameras and from the command interface to control the robot by means of a game pad joystick.

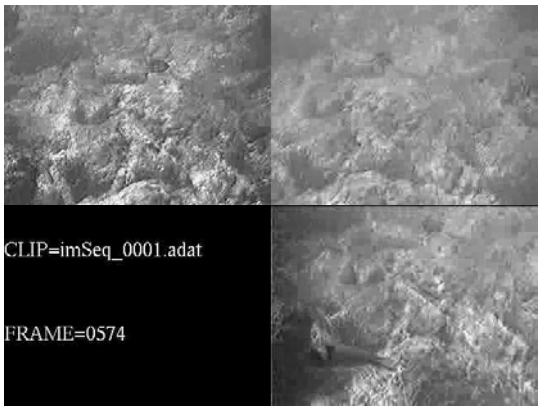
B. *Trinocular Sensor Package*

Due to the inherent physical properties of the marine environment, vision systems for aquatic robots must cope with a host of geometrical distortions: colour distortions, dynamic lighting conditions and suspended particles (known as 'marine snow'). The unique nature of the aquatic environment invalidates many of the assumptions of classic vision algorithms, and solutions to even simple problems -- such as stereo surface recovery in the presence of suspended marine particles -- are not yet known.

A fundamental problem with visual sensing in the aquatic robotic domain is that it is not possible to assume that the sensor only moves when command to. The aquatic medium is



(a) The Sensor



(b) Raw Aquatic footage

Figure 2. The Trinocular sensor package: (a) The sensor shown partially removed from its aquatic housing, the sensor package consists of three firewire CCD cameras, and an IMU. Data from the CCD cameras and the IMU are encoded onto an optical fiber cable and transmitted to the surface via an optical fiber cable. A 12V onboard battery provides power. (b) Raw trinocular data from the sensor (shown here in black and white).

in constant (and in general unpredictable) motion, and this motion complicates already difficult problems in time-varying image understanding. One mechanism to simplify vision processing is to monitor the true motion of the sensor independently of its commanded motion. Inertial navigation systems (INS) have found applications in various autonomous systems for the determination of the relative pose of a vehicle over time. INS make measurements of the physical forces applied to them and thus under normal conditions they provide independent measurements of relative motion. Unfortunately these systems drift, and thus typically they are employed with some secondary sensing system in order to counteract this effect. Here we utilize trinocular vision as this associated sensor. Real time trinocular stereo sensors permit the recovery of 3D surfaces. Integrating an inertial 6DOF navigation system with a trinocular stereo sensor simplifies the registration process by providing relative motion information between frames. With this initial estimate of the camera pose, few features must be used to refine the registration to the global coordinate system.

Figure 2(a) shows the trinocular sensor module and its aquatic housing. The module consists of three Firewire CCD cameras, and an INS. The INS serial signal is converted to a USB signal and is coupled with a pushbutton switch for local control of the device. The combined USB signal and the Firewire signal are converted to an optical signal for transmission via optical fiber to the surface. An onboard 12V battery provides power to the trinocular unit. Figure 2(b) shows raw data obtained with the sensor during recent field trials near Barbados.

C. *Acoustic Sensor Package*

The acoustic localization component consists of arrays of commercially available omni-directional surface-floating hydrophones, whose absolute position can be measured via a combination of GPS, compass, inclinometers and inertial sensors. The underwater sensor unit is equipped with a transducer generating impulsive sound in the audio frequency range. Localization is carried out in two steps, (a) direction of arrival estimation at each array, and (b) estimation of the intersection of the direction lines. Time-delay estimation at each array allows the estimation of direction of arrival at that array [4]. The minimum number of microphones required is three, leading to a system of three linear and one quadratic equation in the coordinates of the direction vector. With more than three microphones, a least mean squares approach is used.

IV. SOFTWARE TASKS

A number of different software tasks of the robot are currently being explored. By combining the best results of these capabilities, a system will be developed that is capable of completing the SASR scenario.

A. *Environmental Modelling*

When the robot is in place making observations, it is often desirable to construct a 3D model of the object being studied. This object may be a pipe that is leaking, or a coral growth that is being monitored. We have developed two complementary (but mutually supportive) methodologies for doing this, one based on stereo and one based on probabilistic extrapolation. In either event, in order to reconstruct a continuous model of the environment, the depth data from the trinocular stereo system must be registered into a global coordinate system since each depth image is independently computed. Current approaches (e.g. [5,6]) to this problem typically use only depth data to minimize an error function for registering multiple point clouds. These approaches are limited

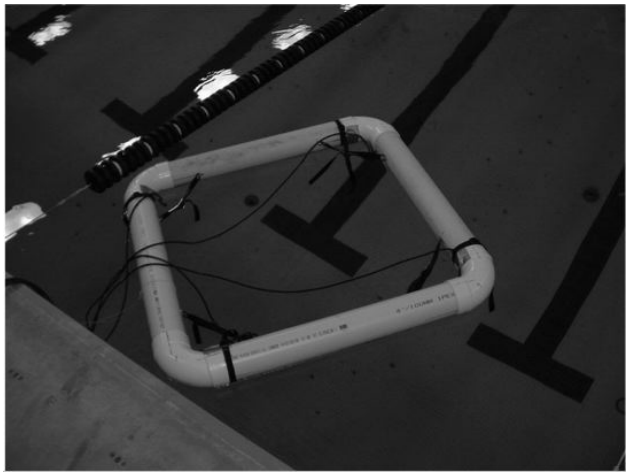


Figure 3. The Acoustic sensor hardware

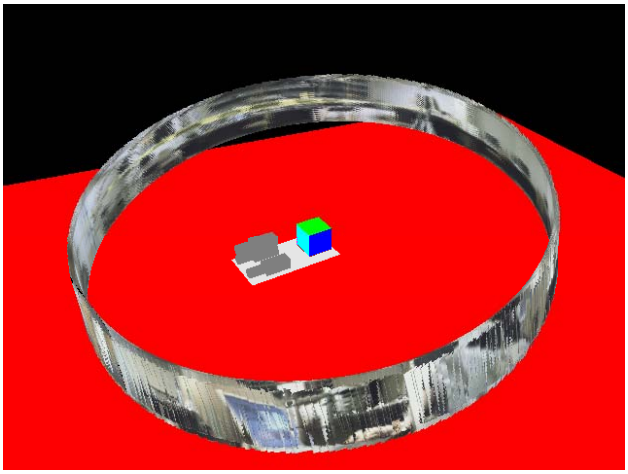


Figure 4. INS-based imagery recovery. The figure shows raw data collected from the sensor in the lab arranged based on the rotational estimate returned by the INS. There is very little drift in terms of orientation, although absolute distance information drifts quite quickly.

by the fact that they rely heavily on small motion between point clouds and thus they overlap. If a large motion were to occur due to underwater currents affecting the sensor package, then these types of algorithms would fail to produce a continuous 3D surface reconstruction of the environment. Using an inertial sensing system, this problem can be alleviated by providing a good initial guess to the registration. When registering the point clouds, the inertial data is used to align the data sets and other Bundle Adjustment [7] and ICP [8] algorithms are used to refine the estimate if an overlap occurs. If there is no overlap, the integrated pose from the inertial data is the best estimate of the sensor's motion. The inertial data can only be trusted for several seconds due to accumulating errors in the integration of the rotational rates and accelerations.

Figure 4 shows the INS in action. Raw video data from the camera is rendered along the edge of a cylinder based on the rotation of the camera as obtained by the INS. The INS maintains a very accurate estimate of orientation although its positional accuracy is quite poor. An integrated process that combines both vision and INS data is required.

B. *Acoustic-based Vehicle Localization*

In order to complete the SASR scenario, a critical capability of the vehicle is to be able to revisit a previously visited position. Maintenance of pose with respect to a global coordinate system is key. Under the AQUA project a number of acoustic and vision-based localization processes are being explored.

Estimation of sound source position as the intersection of direction lines. Suppose that the vehicle is augmented with an acoustic source. The position of this acoustic source can be estimated using an acoustic array mounted at a known position under water. The position of the sound source is estimated through considering multiple lines in 3D space emanating from the reference points of the microphone arrays and along the direction of arrival vectors. The Sound source position is estimated as the intersection of these lines. Computationally, the optimal estimate of the source position is the point that has minimal overall distance from these lines. The overall distance to the unknown source position $P(x,y,z)$ is a quadratic function of the unknowns x,y,z, leading to a linear system of equations in x,y,z that can be solved for using standard techniques.

Signal processing for time delay estimation. In order to calculate reliable time delays between the arrivals of the sound signals, two channels of audio data from two different hydrophones are correlated. Peaks of the correlation function are identified. The location of the peak corresponds to the time-delay estimate. Before correlation, filtering is carried out to reduce noise, and then a signal variance test is performed to check the presence of a sound source [3]. The audio frequency region of interest is (200 - 4000 *Hz*) so as to eliminate high frequency noise as well as the common 60*Hz* electric interference and its second harmonic at 120*Hz*, extracted using a band pass digital FIR filter described in [2]. The variance of a source signal is typically greater than the variance of the background noise. Since the maximum time delay between two microphones can be calculated through the length of the baseline divided by the speed of sound, it would be required that the time delay of the maximum peak of the correlation function is inside a range defined by the maximum delays. This has the effect of reducing the likelihood of false peaks. The final step for the time delay estimation is to cluster the time delays estimated from a number of consecutive and non-overlapping signal time windows [3]. We discard outliers and compute the mean value over the remaining as the final time delay estimate.

Experimental Results: Besides simulations, we have performed experiments in a pool using hydrophones and in the

air using microphones with a similar geometry to that of the pool (scaled to account for the different sound propagation speeds in the two media). In the experiments, we generated impulsive sound by hitting a metal appliance once per second. The problem of designing a transducer for the underwater robot remains to be addressed. The water experiments were carried out in a pool (25.0*m* long, 20.0*m* wide, and 4.3*m* deep at the flat part of the deep end). The listening apparatus consists of four DolphinEar/PRO omni directional hydrophones, which are attached at the corners of a square buoy of size 1.0 x 1.0*m*, shown in Fig. 3. Sampling frequency was 44100 *Hz*, signal resolution 16 bits, sample size 2048 samples.

C. *Vision-based vehicle localization*

An alternative to audio-based localization is to utilize imagery captured from a camera onboard the vehicle to aid in localization. Two approaches are being developed: one based on the use of natural underwater features, and a second based on manually inserted surface-based visual beacons. The use of natural landmarks entails simultaneous localization and mapping (SLAM) while the use of manual landmarks uses beacons with known positions. The beacon-based approach uses an upward-looking camera on the underwater robot to observe a set of (control) points of a known extended rigid buoy at the surface. The buoy itself is localized with respect to a global coordinate system using a combination of GPS, compass, inclinometers and inertial sensors. The problem can be stated formally as follows: Given a set of *m* control points P_i, i=1,2,...*m* whose 3-dimensional coordinates (x_i, y_i, z_i) are known in some global coordinate system, and given an image (taken by a calibrated camera) in which the *m* control points are visible, determine the location (relative to the coordinate system of the control points) of the camera from which the image was obtained.

Let the coordinates of control point P_i in the camera-centered coordinate system be (x_i', y_i', z_i') and let the image coordinates of point i be (u_i, v_i). Assuming perspective projection, we obtain:

$$u_i = \frac{fx_i'}{z_i'} \qquad v_i = \frac{fy_i'}{z_i'}$$

The global coordinate system and the camera coordinate system are related by a rotation R and a translation T. The 3D coordinates of a control point in the two coordinate systems P_i and P_i' respectively are by:

$$P_i' = RP_i + T$$

The goal in camera viewpoint recovery is to determine R and T, knowing the *m* correspondences between image points and control points. The problem of camera viewpoint recovery is solved in two steps. First, the control point coordinates in the camera-centered coordinate system are derived from the camera model and geometric constraints on pairs of points in the two coordinate systems. Second, the transformation between the two coordinate systems is found based on the correspondences between the two coordinate sets of the same points.

Each image point provides two constraints on the three camera-based coordinates for each control point. Furthermore, since (x_i', y_i', z_i') and (x_i, y_i, z_i) represent the same point in 3D space, the distance between the pairs of points is the same in the two coordinate systems, giving three constraints, one for each pair of control points. Thus, with three control points, there are enough constraints to solve for the unknown camera-based coordinates of the control points. With *m* control points, where $m > 3$, we have $2m$ perspective projection equations, and $m(m-1)/2$ pair wise distance constraints, i.e. an over-determined problem. To solve it, we minimize the total squared difference between pair wise distances. If we express x_i', y_i' in terms of z_i', we obtain an unconstrained optimization problem in terms of z_i', where the objective function is a polynomial of degree 4.

After estimating the coordinates of the points P_i in the camera coordinate system, the remaining problem is to identify the transformation between the camera and the global coordinate system. This requires the solution of another least squares optimization problem.

D. *Vision-based Mapping and Localization (VSLAM)*

Our approach to vision-based mapping and localization is based on appearance-based features that are learned from the environment, and is derived from terrestrial techniques examined previously [9]. The process is based on 3 distinct computational processes: detection of potential visual landmarks, landmark matching and tracking, and landmark estimation. The key principle is that landmarks are defined in the appearance domain -- that is from video data -- without an attempt to recover 3D structure. This allows for the use of visual features even in situations where 3D recovery is problematic, and hence provides for a pose estimation mechanism that is complementary to the other processes use to navigate the vehicle. Further, since the landmarks are learned in the underwater environment at hand, the method avoids a dependency on particular predetermined attributes of the environment. Finally, since reliability of each visual feature is explicitly modeled, partial pose estimation case occur even when the features are too impoverished to allow for accurate estimation of all degrees of freedom.

Prototype natural landmarks are detected by an interest operator, which selects regions of the image that should be detectable in subsequent views (several interest operators have been considered in this context [10]). Models for these landmarks are incrementally constructed as a function of both their visual appearance and position, and possible landmarks that prove unstable are discarded. This allows the system to learn domain-specific features for use in localization and to

estimate how useful each landmark is for various aspects of pose estimation and for various positions. Finally, features that are reliably recognized, are modeled statistically and used for pose estimation using a voting scheme that robustly combines the estimates from each landmark (see Fig. 4).

E. Statistical depth recovery from monocular data

In order to build more reliable models of the underwater environment, we are also developing techniques for monocular shape estimation from video data augmented by partial range estimates. The approach is based on extrapolation of the depth map given some initial set of depth estimates. These depth estimates are extrapolated using the video (intensity) data under the assumption that the combination of intensity and depth at each point in an image can be modeled as a Markov Random Field, as described in [11]. That is, given an augmented depth map

$$V = (I, Z)$$

where *I* is an image

$$I = \{i_{x,y}\}$$

and *Z*, a depth map,

$$Z = \{z_{x,y}\}$$

we estimate the probability P() of an augmented depth value from its intensity and neighboring values

$$z_{x,y} = \arg\max(P(z_{x,y} \mid i_{x,y}, v \; in \; N(x,y))$$

where *N(x,y)* is set of augmented depth values in the neighborhood of *(x,y)*. This expresses the dependence of the depth estimate at a point of the neighboring depth and intensity values, and on the intensity at that point. This probabilistic dependence is precisely the transition function of the Markov random field.

Given that this is the case, the transition function for the MRF is computed from whatever partial data is available initially (for example as extracted from the stereo head). These transition probabilities are then used to compute the depth component of the image where measurements are absent, using the intensity component as a constraint. Conceptually, the approach has some relationship to shape-from-shading although without any dependence on an a prior knowledge of the scene reflectance function, nor on a uniform albedo. While such extrapolation from either depth alone, or from intensity alone, appears to be exceedingly difficult, the combination of intensity data with constraints from sparse depth estimates seems to make the problem tractable. Preliminary tests in terrestrial environments suggest that the approach is effective, although validation in the target environment remains to be carried out.

In recent sea trials of the robot, we have observed that this technique may be useful not only for monocular depth recovery, but also for deblurring and removal of scattering artifacts.

The financial support of the IRIS NCE Program and NSERC Canada is gratefully acknowledged.

V. Discussion and Future Work

In recent sea trials, the physical robot, trinocular vision system and other components were tested in the Caribbean Sea up to a depth of about 23 feet. Once the buoyancy was manually adjusted to compensate for the salinity where the test was conducted, the robot performed well using nearly-neutral buoyancy. Gait control was accomplished manually but controlling the robot using only the forward-mounted cameras proved to be a challenge. In ongoing work we will be adding both an inclinometer readout and tele-robotic control modes to improve manual controllability. In addition, we are also developing automated control modes.

One of the key challenges of the project is the extension of the SLAM philosophy (Simultaneous Localization and Mapping) into motion in three dimensions, with robot pose depending on six degrees of freedom. Much of the SLAM research so far has been restricted to two-dimensional manifolds, either planar or topographic surfaces, with robot pose depending on three degrees of freedom. Odometry information of the same nature as in terrestrial robots is difficult to obtain in the underwater domain, so one has to rely instead on an accurate dynamic model of the underwater robot combined with inertial sensors and sensors of external fields (gravity, earth's magnetic field) to come up with differential position estimates for mapping. Furthermore, GPS information, which is available to outdoor robots, is not available underwater, so the absolute position of the underwater robot needs to be constrained by its relative position with respect to surface vessels with access to GPS signals. The key sensing modality for mapping in this project is vision, aiming towards smaller scale mapping than that based on sidescan sonar.

Acknowledgments

The technical assistance of Jeff Laurence and Matt Robinson is gratefully acknowledged.

References

[1] R. Altendorfer, N. Moore, H. Komsuoglu, M. Buehler, H. B. Brown Jr., D. McMordie, U. Saranli, R.J. Full, D.E. Koditschek. "RHex: A Biologically Inspired Hexapod Runner." Autonomous Robots 11:207-213, 2001

[2] B. Kapralos and M. Jenkin and E. Milios. Audio-visual localization of multiple speakers in a video teleconferencing setting. *International Journal of Imaging Systems and Technology* , 13(1):95-105, 2003.

[3] B.T. Luke. *Agglomerative Clustering*. http://fconyx.ncifcrf.gov/~lukeb/agclust.html, last accessed on July 8, 2003.

[4] G. L. Reid and E. Milios. Active stereo sound localization. *The Journal of the Acoustical Society of America*. 113:185-193, January 2003.

[5] A. J. Stoddart, A. Hilton, Registration of multiple point sets, Proc. 13th Int. Conf. on Pattern Recognition, pp B40-44 Vienna, Austria, (1996).

[6] Chu-Song Chen and Yi-Ping Hung and Jen-Bo Cheng. "A New Approach to Fast Automatic Registration of Partially Overlapping Range Images." IEEE Transactions on Pattern Analysis and Machine Intelligence, 21(11):1229-1234, 1999.

[7] Bill Triggs and Philip McLauchlan and Richard Hartley and Andrew Fitzgibbon. Vision Algorithms: Theory and Practice. Triggs, W. and Zisserman, A. and Szeliski, R. Springer Verlag, 2000, p. 298-375.

[8] Gregory C. Sharp and Sang W. Lee and David K. Wehe. "ICP Registration using Invariant Features" IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(1): 90-102, 2002.

[9] R. Sim and G. Dudek, "Learning environmental features for pose estimation", Image and Vision Computing, vol 19, num 11, pp 733-739, Elsevier Press, 2001. 733-739, 2001.

[10] Robert Sim and Gregory Dudek, "Learning Generative Models of Scene Features", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Hawaii, 2001, 7 pages.

[11] Torres-Mendez, Luz-Abril, and Gregory Dudek, "Range Synthesis for 3D Environment Modeling", Proceedings of the IEEE/RSJ Conference on Intelligent Robots and Systems (IROS), Las Vegas, NV, Oct. 2003, 8 pages.