

A visually guided swimming robot

Gregory Dudek^{††}, Michael Jenkin^{**}, Chris Prahacs^{*}, Andrew Hogue^{**}, Junaed Sattar^{††},
Philippe Giguere^{††}, Andrew German^{**}, Hui Liu[‡], Shane Saunderson^{*},
Arlene Ripsman^{**}, Saul Simhon^{††}, Luz-Abril Torres^{††}, Evangelos Milios[‡], Pifu Zhang[‡], Ioannis Rekleitis^{††}

^{††}School of Computer
Science
McGill University
3480 University St.
Montreal, PQ, Canada

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

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

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

Abstract – We describe recent results obtained with AQUA, a mobile robot capable of swimming, walking and amphibious operation. Designed to rely primarily on visual sensors, the AQUA robot uses vision to navigate underwater using servo-based guidance, and also to obtain high-resolution range scans of its local environment. This paper describes some of the pragmatic and logistic obstacles encountered, and provides an overview of some of the basic capabilities of the vehicle and its associated sensors. Moreover, this paper presents the first ever amphibious transition from walking to swimming.

Keywords—autonomous robot, aquatic robot, robotic sensing

I. INTRODUCTION

The aquatic environment presents an almost ideal test-bed for the evaluation and development of robotic technologies. The environment is highly dynamic and three dimensional. Vehicles operating in this realm must cope with unpredictable 3D motion of the vehicle itself complicating any task to be performed. Extremely limited off-board communication underwater requires that robots must operate fully autonomously or under operator control through a tether. Many surveillance and data collection tasks require that a vehicle be positioned in the same location for a period of time. Surge and underwater currents prevent this from happening, thus robust and accurate pose maintenance is a requirement for aquatic vehicles. Also, standard mapping techniques originally developed for indoor robot navigation are not general enough to describe the inherently six degree of freedom aquatic environment. A natural choice for sensing is to use cameras and computer vision techniques to aid in navigation and trajectory reconstruction. However, a host of other problems plague underwater computer vision techniques. For example, the turbidity of the water caused by floating sedimentation (‘aquatic snow’) and other floating debris wreaks havoc on standard techniques for visual understanding of the world. The environment itself provides unique challenges for the mechanical design of vehicles – keeping the water away from electrical equipment is essential and rarely required for navigating university laboratory hallways.

Given the complexity of the aquatic domain, and the desirability of autonomous and semi-autonomous systems to aid in its exploration, there has been a long history of

development of aquatic robots (cf. [2]) and their sensors (cf. [15]). The vast majority of aquatic vehicles are controlled by thrusters and control surfaces. Many are designed to be supported by a sizeable shore- or ship-mounted infrastructure to provide external power, communication and tether due to their large size.

In contrast to many of these earlier aquatic vehicles, within the AQUA project we are developing a small-scale legged swimming robot whose core sensing modality is image-based; such a robot is small, portable and maneuverable. The specific target application we are working towards is the environmental assessment and longitudinal analysis of coral reef environments. These environments are difficult to navigate with traditional autonomous underwater vehicles or ROVs (remotely operated vehicles) due to the shallow depths and the significant presence of marine life. A major drawback for traditional vehicles is the usage of thrusters at shallow depths which stirs up much of the ocean floor reducing visibility and endangering the coral and marine life.

Reefs are ecologically important environments that are seriously threatened yet a clear objective assessment of their health is a serious challenge to which automated methods may prove valuable. In particular, coral reefs occupy only 0.7% of the ocean floor, but provide homes and vital nursery grounds for 25% of all marine species on the planet. Furthermore, reef monitoring serves as a valuable template for a range of other significant applications. A well established bio-assessment methodology is to either repeatedly visit a set of selected locations or to regularly swim over a particular trajectory and visually inspect the marine flora and fauna. This specific task is labor intensive for human operators and, as a sample application, it may be well suited to the AQUA vehicle.

This paper provides some details of the design and control of the AQUA vehicle and of one of its primary visual components, a combined inertial-trinocular camera module. Further details of the vehicle, and in particular details of an acoustic localization system for the robot can be found in [4].

II. THE AQUA PLATFORM

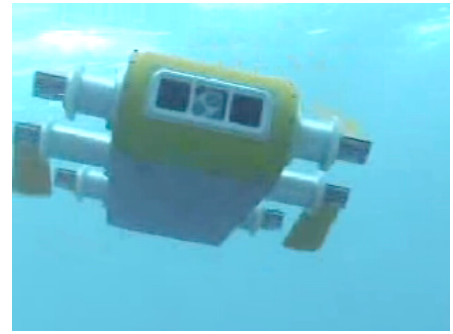
The AQUA vehicle represents a novel approach to aquatic vehicle design. Rather than relying on thrusters and control surfaces, the AQUA vehicle relies on legged locomotion. The



(a) On the beach



(b) Surface swimming



(c) Free swimming

Figure 1. The AQUA robot. (a) The vehicle with its amphibious legs walking on the beach. (b) Shows the vehicle swimming on the surface of the ocean using its amphibious legs. (c) Shows the vehicle swimming in the open ocean using its fins. A number of different leg structures have been tested on the vehicle. Two different sets of legs are shown here.

vehicle uses its six actuators, coupled to specially designed legs and/or fins, to move either as a walking vehicle or as a legged swimmer. Legs provide the vehicle with a variety of different locomotive strategies including the ability to walk on solid surfaces, and to swim on the surface of the water or in the open ocean. Walking behaviors are based on variations of a rotary gait (Figure 1(a)), while swimming behaviors depend on variations of an oscillating leg motion (Figure 1(b)).

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]. AQUA's required capabilities include surface and underwater swimming, diving to a depth of 15m, station keeping and crawling at the bottom of the sea. For propulsion, the vehicle deviates from traditional ROV through the use of 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: surge, heave, pitch, roll and yaw. An inclinometer and a compass onboard are used to assist 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 is encased within an aluminum waterproof shell and displaces about 18 kg of water. The robot is powered by two onboard NiMH batteries providing over two hours of continuous operation. Onboard computation is performed using PC/104 computers, one running a real-time operating system (QNX), and the other an embedded version of Linux. Camera, sensor and control signals are sent to a floating platform at the surface via fiber-optic tether. This information is used by the operator to control the robot via usage of a gamepad controlled command interface.

Each of AQUA's six legs is individually controlled by a single axis motor. The onboard computer provides real-time control of the set of six limbs. Although a number of different leg designs have been developed and tested on the vehicle, a common design feature of the legs is that they are compliant.

The spring energy stored in the legs as they bend under load is an integral part of the vehicle's locomotion strategy.

AQUA's unique locomotion strategy provides great flexibility in terms of potential modes of locomotion. Figure 1 shows the AQUA vehicle moving in three very different modes. In Figure 1(a), the AQUA vehicle is shown with its amphibious legs walking along the beach near the surf zone. During walking locomotion, the AQUA vehicle uses its six single-jointed limbs to walk using an alternating gait based upon a gait originally designed for the Rugged RHEX platform. When the vehicle moves through the surf zone, it transits to a surface swimming gait (see Figure 1(b)). Here the amphibious legs are used to provide thrust along the surface of the water. Finally, when the water depth is sufficient, the vehicle can transit to an open water swimming mode in which the vehicle "swims" in the full 6DOF space that exists in the open ocean. Figure 1 also illustrates that the AQUA vehicle can be deployed with different sets of legs. Individual legs are designed to be easily attached and removed from the vehicle, and the design of a robust set of 'all weather' legs is the topic of ongoing research.

Limbed locomotion requires the development of appropriate gaits to drive the legs in a coordinated manner in order to move the vehicle. A set of gaits have been developed that provide controlled motion of the vehicle when walking, swimming on the surface and swimming in open water.

AQUA is a 6DOF vehicle. Each leg has a single controllable degree of freedom used for gait generation. Figure 2 shows a sequence of snapshots of AQUA's walking gait. The sequence of snapshots should be viewed from left to right and top to bottom. The walking gait is a basic hexapod walking gait. Swimming in open water permits a rich class of alternative gaits and behaviors, and allows motion with six degrees of freedom (although there is often coupling). Currently there are two mode of operation for AQUA. First, the most commonly used teleoperation, where an operator can send commands to the robot via the optical cable to control the motion. Second, currently under development are a set of prearranged behaviours based on visual servoing for following



Figure 2. Eight snapshots of the AQUA vehicle walking along the beach. The individual frames should be viewed in a left to right, top to bottom order. AQUA maintains a statically stable gait while walking.

a visual target and also for following a set trajectory, In particular, by application of the appropriate gait the robot is able to roll, pitch, yaw, surge, and heave. At present these behaviors have been tested under manual control and the robot appears rather easy to control manually. Complex 6DOF trajectories can be challenging to track, and this is one of the factors motivating the use of servo-controlled gaits, even for manual vehicle guidance.

III. SENSING WITH AQUA

The AQUA robot enclosure contains both forward- and rear-facing video cameras. These sensors are used primarily for vehicle teleoperation and visual servoing tasks. Other sensors (e.g. the acoustic localization sensor and the trinocular sensor) are currently being evaluated in external housings and will be integrated into the body of the robot in later versions of the vehicle. The acoustic localization sensor is described in some detail elsewhere ([4]). Here we provide some details on the trinocular vision sensor system, color correction processing and video servoing.

The AQUA platform is intended for use as a device to aid in environmental assessment, and in particular to aid in the assessment of reef damage and rehabilitation. Thus two key inspection tasks are the collection of video footage, and the construction of 3D surface and volumetric models of the reef structure off-line. Later, off-line comparison of these models over time can be used to assess the rate of variation, damage or (hypothetically) restoration of the environment. In the context of the AQUA project we are addressing this task as a general Site Acquisition and Scene Re-inspection (SASR) task. A typical scenario in a SASR task is as follows. The 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. The robot then surveys the environmental location, possibly in a

teleoperation mode, and a surface model of the reef is recovered. Once measurements are made, the robot then returns home autonomously. At a later date, the robot actively guides – and potentially controls – its motion to the previously visited site in order to collect additional data.

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 fully understood.

Color Correction

For many inspection and observation tasks, high quality image data is desirable. We have developed a technique for image enhancement based on training from examples. This allows the system to adapt the image restoration algorithm to the current environmental conditions and also to the task requirements. Image restoration involves the removal of some known degradation in an image. Traditionally, the most common sources of degradation are due to imperfections of the sensors, or in transmission. For the case of underwater images, additional factors include poor visibility (even in the cleanest water), ambient light, and frequency-dependent scattering and absorption both between the camera and the environment, and also between the light source (the sun) and the local environment (i.e. this varies with both depth and local water conditions). The result is an image that appears bluish, blurry and out of focus. Most prior work tends to approximate the deblurring and noise processes by idealized mathematical models. Such approaches are often elegant, but may not be well suited to the particular phenomena in any specific real environment. Image restoration is difficult since it



Figure 3: Uncorrected and corrected images. The color correction and deblurring is accomplished using a learning based Markov Random Field model.

is an ill-posed problem: there is not enough information in the degraded image alone to determine the original image without ambiguity.

As proposed in [11], our approach is based on learning the statistical relationships between image pairs. In our case, these pairs are the images we actually observe and corresponding color-corrected and deblurred images. Our approach uses a Markov Random Field model to learn the statistics from the training pairs. This model uses multi-scale representations of the corrected (enhanced) and original images to construct a probabilistic enhancement algorithm that improves the observed video. This improvement is based on a combination of color matching, correspondence with training data, and local context via belief propagation, all embodied in the

There are a number of factors that influence the quality of the results. These include the adequate amount of reliable information as an input and the statistical consistency of the images in the training sets.

Visual Servo Control Subsystem

The visual servo-control subsystem is used to permit semi-autonomous navigation. At present, this is aimed at allowing the robot to follow a human guide during the mapping or image acquisition process. While not currently implemented, this should allow the robot to subsequently re-play the same trajectory in fully autonomous mode. Servo-control also has several other task-specific applications. While the aquatic medium essentially provides low-pass filtering for the vehicle dynamics, the servo-control system still needs to operate at near-real time. This requirement is even more stringent for terrestrial behaviors. Thus, a second processor is devoted exclusively to vision processing. This is a Pentium M processor running at 1.1 GHz with CPU frequency scaling capability, a Gigabyte of RAM and conforms to the PC/104 Plus form factor. While faster processors are available, low power consumptions and a desire to passive cooling and the associated physical robustness it entails make this a suitable choice. The vision stack uses Linux and runs off a 512 Megabyte CompactFlash card. For image acquisition a IEEE1394 digital color camera with resolution of 640-by-480 pixels is used with a PC/104 plus IEEE1394 capture card. The vision processor interfaces with the camera through the IEEE1394 capture card. The tracker has been written in C++ using an open source vision library called VXL (Vision-Something-Libraries – the X stands for “something”). VXL includes libraries for computer vision, mathematics, file i/o, image processing and other useful vision processing functionality. Communication between the control and vision stack takes place over the ethernet using the UDP protocol.

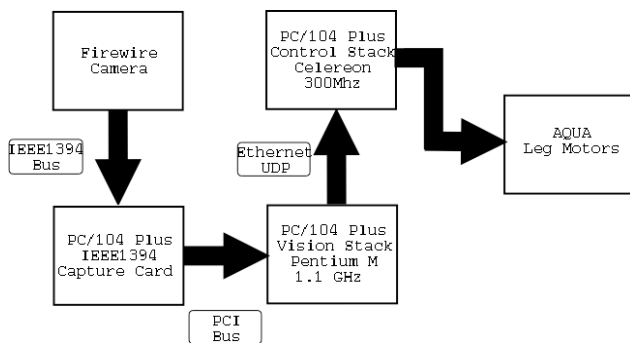
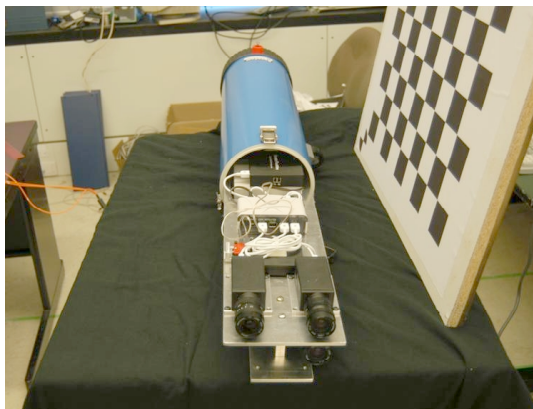


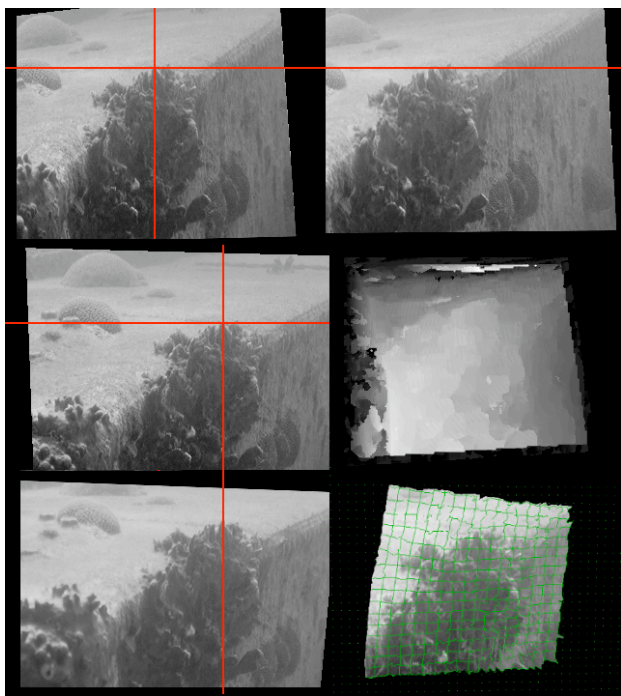
Figure 4: The AQUA Visual Servoing Architecture

Markov random field. Training images are small patches of regions of interest that capture the maximum of the intensity variations from the image to be restored. The corresponding pairs, i.e. the ground truth data containing the restored information from the same regions, are captured when lights mounted on the robot are turned on. Some experimental results are shown in Figure 3.

Trinocular Vision



(a) The Trinocular-IMU Sensor



(b) Rectified Imagery, Disparity Map and Reconstructed 3D Mesh of Coral growing on a sunken barge.

Figure 5. 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) The recovered 3D surface and the rectified imagery that was used to obtain the recovered surface.

A fundamental problem with visual sensing in the aquatic domain is that it is not possible to assume that the sensor only moves when commanded to. The aquatic medium is 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 measurement units (IMUs) have found applications in various autonomous systems for the determination of the relative pose of a vehicle over time. IMU devices take measurements of the physical forces applied to them using MEMS gyroscopes and accelerometers and thus under normal conditions they provide independent source-less measurements of relative motion. Unfortunately, due to the use of numerical integration and sensor/temperature characteristics, these systems drift. Typically IMUs 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 5(a) shows the trinocular sensor module and its aquatic housing. The module consists of three Firewire CCD cameras, and an IMU. The IMU signal and the Firewire camera signal are transmitted via optical fiber to the surface. An onboard 12V battery provides power to the trinocular unit. Figure 5(b) shows raw data obtained with the sensor during recent field trials at a coral reef off Barbados. In Figure 5(b) the data has been rectified in order to aid in stereo-based depth recovery.

The trinocular sensor has two operational modes. The first allows the cameras to operate at 640x480 grayscale resolution at 30fps, while the second collects images at 640x480 RGB colour images at 15fps. The drop in frame rate is due to bandwidth limitations on the 1394a bus.

Prior to operation the intrinsic and extrinsic camera parameters are estimated using Zhang's camera calibration algorithm[16] using an underwater calibration target. The appropriate homographies are then applied to each image, rectifying the left-right camera pair to ensure horizontal epipolar lines and the left-bottom pair is rectified to ensure vertical epipolar lines.

A coarse-to-fine trinocular stereo algorithm based on [Mulligan] is used to obtain dense 3D depth maps for each time step. Figure 3(b) shows typical results from the trinocular stereo algorithm. As the 3D structure of individual frames they are integrated together using invariant image features and the 6DOF IMU data.

IV. DISCUSSION AND FUTURE WORK

In recent sea trials, the physical robot, trinocular vision system and other components were tested in the Caribbean Sea near Holetown, Barbados up to a depth of about 45 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 teleoperating 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 more sophisticated telerobotic control modes to improve manual controllability.

The vehicle was able to exhibit the full range of swimming motions, as well as the previously-tested walking behaviors demonstrated on the RHex family of robots. Swimming under the control of visual servo control was demonstrated to be an effective control paradigm even in water with limited visibility. Initial tests indicated that the vehicle was able to follow a cooperative free-swimming driver. The tradeoff between proximity to a diver, safety, and robustness is an area for future work. In addition, in preliminary tests of amphibious operations the vehicle was able to make the transition from walking to swimming and vice-versa, although this entails skilled manual control and many technical issues remain to be considered.

The trinocular sensor package was tested on both naturally occurring and man-made underwater structures (wrecks and other structures) at different depths and under different environmental conditions. Aquatic snow and other particulate matter provide challenges for traditional stereo algorithms, and work is currently underway to investigate how to detect and reject invalid matches that occur due to these effects.

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 three-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

We would like to thank Jeff Laurence, Matt Robinson, Jim Zacher, Michelle Theberges, Dimitri Marinakis, and Sidarth Verma. This project was supported by NSERC and by NCE IRIS.

REFERENCES

- [1] Altendorfer, R., Moore, N., Komsuoglu, H., Buehler, M., Brown Jr., H. B., McMordie, D., Saranli, U., Full, R. J. and Koditschek, D. E. "RHex: A Biologically Inspired Hexapod Runner." *Autonomous Robots* 11:207-213, 2001
- [2] Balasuiya, A. and Ura, T., Underwater cable following by twin-burger 2, Proc. IEEE ICRA, Korea, 2001.
- [3] Chen, C.-S., Hung, Y.-P. and Cheng, J.-B.. "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.
- [4] Georgiadis, C., German, A., Hogue, A., Liu, H., Prahacs, C., Ripsman, A., Sim, R., Torres, L.A., Zhang, P., Buehler, M., Dudek, G., Jenkin, M. and Miliotis, E. "AQUA: An aquatic walking robot." *Proceedings of the IEEE/RSJ/GI International Conference on Intelligent Robots and Systems (IROS)*, Sendai, Japan, 2004.
- [5] Kapralos, B., Jenkin, M. and Miliotis, E.. Audio-visual localization of multiple speakers in a video teleconferencing setting. *International Journal of Imaging Systems and Technology*, 13(1):95-105, 2003.
- [6] Luke, B. T. *Agglomerative Clustering*. <http://fconyx.ncifcrf.gov/~luke/agclust.html>, last accessed on July 8, 2003.
- [7] Reid, G. L. and Miliotis, E. Active stereo sound localization. *The Journal of the Acoustical Society of America*. 113:185-193, January 2003.
- [8] Sharp, G. C., Lee, S. W. and Wehe, D. K. "ICP Registration using Invariant Features" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(1): 90-102, 2002.
- [9] Sim, R. 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] Sim, R. and Dudek, G., "Learning Generative Models of Scene Features", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Hawaii, 2001, 7 pages.
- [11] Singh, B., Freeman, W. T. and Brainard, D. Exploiting spatial and spectral image regularities for color constancy. *Workshop on Statistical and Computational Theories of Vision*. Nice, France, October, 2003.
- [12] Stoddart, A. and Hilton, A., Registration of multiple point sets, Proc. 13th Int. Conf. on Pattern Recognition, pp B40-44 Vienna, Austria, (1996).
- [13] Torres-Mendez, L.-A., and Dudek, G., "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.
- [14] Triggs, B., McLauchlan, P., Hartley, R. and Fitzgibbon, A. *Vision Algorithms: Theory and Practice*. Triggs, W., Zisserman, A. and Szeliski, R. (Eds.) Springer Verlag, 2000, p. 298-375.
- [15] Xu, X. and Negahdaripour, S., Vision-based motion sensing for underwater navigation and mosaicing of ocean floor images. Proc. IEEE Oceans'97, 1412-1417, 1997, Halifax, NS.
- [16] Zhang, Z. A Flexible new technique for camera calibration. Technical Report MSR-TR-98-71, Microsoft Research, 1998.