An Operational Real-Time Large-Scale Visual Mosaicking and Navigation System

Kristof Richmond* and Stephen M. Rock*†
*Aerospace Robotics Laboratory, Stanford University
Durand Bldg, Rm 028
496 Lomita Mall
Stanford, CA 94305-4035
Email: {kristof,rock}@stanford.edu

†Monterey Bay Aquarium Research Institute
7700 Sandholt Road
Moss Landing, CA 95039-9644

***Abstract*— One use of ROVs (and potentially AUVs) is the generation of visual mosaics of areas of the ocean floor. A visual mosaic is formed by joining together multiple images taken by the vehicle as it surveys the sea floor to form a larger, composite view. In order to cover an area completely while guaranteeing overlap between images, precise measurement and control of the vehicle is required. External arrays such as LBLs can provide the precise navigation information required, at considerable expense in deployment and with limited operational range. Self-contained systems using DVLs and IMUs, do not suffer from these limitations but are subject to dead-reckoning drift. However, the imaging system itself can provide a means to eliminate drift. As long as the vision system can see a previously-visited area, the position measurement error relative to that part of the environment is bounded, and does not depend on vehicle path length or integration time. Additionally, the position information can be used to display a real-time mosaic to the user, which grows as the vehicle moves. Vision is subject to outages which do not affect a dead-reckoning sensor, such as dust clouds obscuring the view, and it is limited in its range from bottom. However, fusing the measurements of these two self-contained, complementary sensors—dead-reckoning and vision—provides a robust sensor which can provide the precision to guarantee complete area coverage and sufficient overlap in large-scale mosaics without the need to deploy an external positioning array.**

This paper presents an online seafloor mosaicking and navigation system which exploits the complementarity of dead-reckoning with a DVL and direct environment-relative sensing using vision. The mosaicking and navigation system operates in real time, and the calculated position provides a measurement which can be fed back to control the vehicle position relative to the environment and to display a navigation-grade mosaic to the user in real time. The results of field trials conducted in Monterey Bay using the MBARI ROVs *Ventana* and *Tiburon* are presented.

Fig. 1. Navigation-grade mosaic of a whalefall at the bottom of Monterey bay taken from ROV *Tiburon*. Using the real-time visual mosaicking and navigation system for control, the ROV performed a lawnmower pattern at an altitude between 2.1 m and 3.4 m. The maximum height of the skeleton above the sea floor was about 0.9 m. The mosaic covers an area of about 7 m $\times$ 11 m and contains 120 tiles.

I. Introduction

The Stanford Aerospace Robotics Lab (ARL), in cooperation with the Monterey Bay Aquarium Research Institute (MBARI), has developed and demonstrated a real-time visual mosaicking and navigation system for use as a pilot aid on ROVs. The system provides high-precision, environment-relative vehicle positioning and control without the use of external positioning arrays and provides the user with an evolving view of the area being explored by the vehicle along with the capability to control the vehicle relative to the environment. The goal is to provide a self-contained real-time navigation system enabling direct interaction with objects in the local environment through an intuitive, information-laden interface.

Visual mosaicking—stitching together snapshots to produce larger composite images—is a powerful tool for benthic exploration [1]. The limited propagation of light in water constrains the possible size of single-frame images that can be acquired. In order to "remove the water" and gain a large-scale overview of the seafloor, a composite mosaic image must be constructed. Historically, these mosaics have been assembled manually the vehicle capturing the images has returned to the surface. More recently, the process has taken advantage of the advances in digital computing and has become much more automated.

In order to provide a complete overview of the area, the entire surface must be photographed. That is, all images (also referred to as tiles) in the mosaic must overlap with their neighbors, both along the vehicle track and side-to-side. This places constraints on both the frequency with which images must be taken along-track, and the distance between successive vehicle tracks. This in turn places additional burdens on the vehicle system by requiring accurate measurement and control of the vehicle position to know both where images were taken and to ensure full coverage.

To provide navigation information of sufficient precision to an ROV or AUV, an external transducer array, such as an LBL system, can be employed. While they provide accurate navigation information, transducer arrays incur additional costs in deploying and calibrating, and once deployed, provide a limited operational area in which the vehicle can navigate with accuracy. In addition, an external array does not provide a direct measurement of position relative to the environment, but only to the array transducers, so that array motion and differencing measurements to obtain relative positions introduce additional errors.

In contrast, self-contained navigation systems—requiring no systems external to the vehicle—have the advantages of substantial operational cost savings, as well as the improved vehicle autonomy which comes from removing any constraint on the area of operation. However, current self-contained navigation solutions, such as IMUs or DVLs, are based on integration of sensed time-derivatives of vehicle position, and thus suffer from dead-reckoning drift in the measurement of vehicle position. The 1-2% of distance travelled drift typical of DVL navigation solutions severely limits the size of visual mosaics for which complete coverage can be guaranteed. Lower drift can be achieved by fusing IMU measurements, though any significant reduction requires high-cost IMUs, and this can only reduce the rate of drift, not bound it.

However, the imaging system can be used not only to gather data for a post-processed mosaic, but also to provide a real-time position measurement. If there are no crossover points in the vehicle trajectory, vision acts as an odometry sensor, which drifts at approximately the same rate as the DVL. However, if portions of the trajectory are close enough together that the vision system is able register current mosaic tiles to ones visited previously, the relative error between the two tiles can be bounded, regardless of the time or vehicle trajectory separating them. Thus, vision is able to eliminate the drift in relative tile positions. These updated measurements can additionally be used to display to the user an on-line mosaic, giving immediate feedback as to the larger vehicle environment.

A vision system has other limitations not suffered by a dead-reckoning sensor, namely a restricted range from the seafloor and potential dropouts due to obscuring dust or marine life. Thus a dead-reckoning sensor—with continuous measurements subject to drift—and a vision system—with potential dropouts but a bounded-error, environment-relative measurement—provide complimentary position sensors which can be fused in a self-contained system which is able to guarantee coverage of an area of interest without needing to resort to expensive external positioning systems.

The system described in this paper takes advantage of the complementarity of these two sensors to create a robust, real-time estimate of the vehicle position along with a visual mosaic of the area visited by the vehicle. The DVL provides a continuous position estimate, even through periods of vision outage. The vision subsystem knocks eliminates drift in the dead-reckoned measurements by taking advantage of cross-over points in the vehicle trajectory, bounding the absolute position error. Instead of producing a fully-optimized mosaic, the focus in the system is on providing real-time feedback on the position of the vehicle and on the evolving view of the environment, sacrificing some of the accuracy of a batch-processed mosaic.

Several components of this system enable this real-time capability. First, displacements between mosaic tiles are quickly computed by a fast correlation method (SLOG + xor) which takes advantage of orientation sensors (pitch, roll, compass) and an altimeter to reduce the dimension of the space over which it searches. Second, in order to overcome periods of vision outage, the system fuses information from dead-reckoned navigation, eliminating the need for a large search for matching tiles when vision is restored. Third, the relative error between new tiles and their neighbors is kept low by taking advantage of vehicle trajectories, such as a lawnmower pattern, which maintain large amounts of overlap between successive passes. High confidence in the relative positions of tiles reduces the number of neighboring tiles which must be tested for matches, generally to just a single tile. Finally, the positions of tiles in the mosaic are computed by a fast, flexible information filter which provides real-time estimates of tile locations for mosaics with thousands of tiles.

These tiles can be displayed in real time at the computed positions to provide live updates of the mosaic progress and the vehicle position. Due to the constraints of real-time computation, lighting variations and distortion of the images are not compensated for and no edge blending is performed, so that the real-time mosaic presents a rough composite image. However, the tiles and measurements provide an excellent initial estimate to feed offline batch processes optimizing over many more parameters and performing blending and smoothing operations to obtain a more visually correct picture. The real-time position feedback and real-time mosaic display allows this system to guarantee visual coverage of a survey area without the need to resort to expensive external absolute position measurements. Figure 1 shows an example of a real-time mosaic produced by this system.

II. BACKGROUND

The use of vision as a position sensor on the sea floor was pursued early on by Marks and Fleischer [2], [3]. The high-speed algorithms they developed form the core of the system described in this paper. Their work has also been extended by Garcia, et. al. into a Kalman filter framework demonstrated in a test tank [4]. Eustice, et. al. [5] have developed methods using image feature point matching to update dead-reckoned navigation information. Additional work in using vision as the *sole* sensor to measure the complete vehicle state has been pursued by Gracias, et. al. [6] and Negahdaripour, et. al. [7].

At the core of the real-time mosaicking and navigation system is an image registration subsystem which quickly provides a measurement of the relative displacement of two images taken by the vehicle camera. The two images are first filtered using the signum of Laplacian of Gaussian (SLOG) operation [8], which extracts textures in the images. The visual offset is then calculated via a fast sum-of-xor correlation. Only two components of the motion are calculated, namely the translation in the plane parallel to the bottom. The remaining components of the motion are taken from other sensors. Constraining the potential motion to two dimensions greatly increases the speed of the correlation computation.

The location of the 2D correlation peak gives the translational offset, and its magnitude is a direct function of the variance in the offset measurement. This maximum correlation value is referred to as the correlation confidence. It is thresholded at an empirically-determined value to determine whether the system has visual lock [9].

With this system, image registration can be performed at frame rate (30 Hz), and provides the basis for a visual odometry system which calculates displacement along the vehicle track, and determines when new images from the live video stream should be stored and added to the real-time mosaic. To do this, live images from the camera are continuously compared to a previously-obtained reference image. When the live image moves far enough from the center of the reference image, it is automatically added to the mosaic and becomes the new reference image. The first image seen when the system is initialized is the initial reference and is taken as the origin of the mosaic coordinate system.

One limitation of vision-based underwater navigation systems is that they suffer from drop-outs. During practical operations, it is not always possible to maintain visual lock on the sea floor. Visual occlusions, such as dust clouds, can come into the field of view, or there may be muddy or sandy areas without sufficient visual variation. In addition, in order to see the bottom consistently, the vehicle is required to operate at a limited altitude (~2 m for *Ventana*). However, mission parameters may require periods of time out of this visual range of the sea floor, necessitating a vision outage. Such outages render the vision component useless for their duration. However, they do not happen silently but rather are indicated by low correlation and vision loss-of-lock.

To overcome possible outages in the vision system, navigation information from a DVL was fused in, providing continuous coverage during periods of vision loss-of-lock [10].

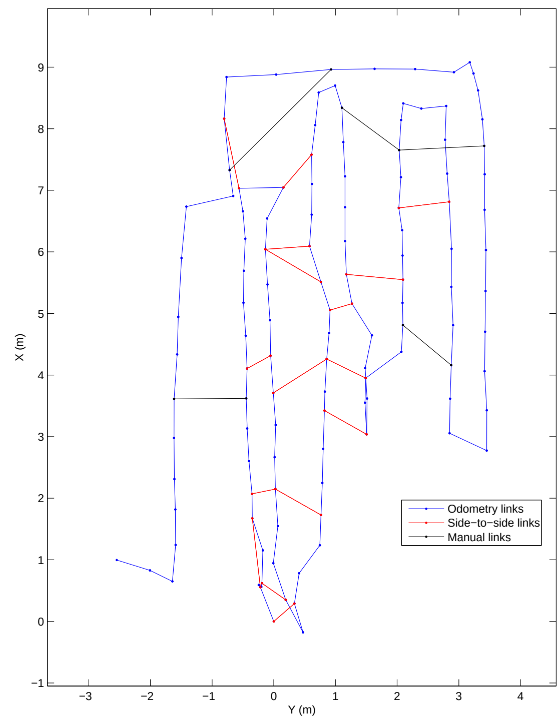


Fig. 2. Graph of links between tiles in mosaic of Figure 1. The initial tile was placed at $(0, 0)$, after which the vehicle flew the trajectory in blue. Side-to-side registration, as performed by the algorithms of Section IV, is shown by the links in red. In the areas of high relative relief, operator intervention to match mosaic tiles was sometimes required.

III. SIDE-TO-SIDE REGISTRATION

While visual odometry provides positioning information and a real-time visualization of the environment the vehicle has visited, it still suffers from errors unbounded in time. Errors in registration between along-track tiles add up just as dead-reckoning errors build up as position derivatives are integrated. In practice, the drift rate of the visual odometry system described here is ~2% of distance traveled (DT) [10], comparable to DVL navigation drift [11], [12].

To perform detailed surveys of any reasonable size, such drift rates are prohibitive. For example, consider a mosaicking survey where the vehicle follows a lawnmower-type pattern, maintaining an altitude such that images it captures cover an area of 2 m $\times$ 2 m. If the minimum image overlap to allow for registration is a (liberally small) 25%, while the maximum desired overlap to allow the survey to cover the area of interest in a timely manner is 75%, then the vehicle trajectory error must be confined to within ± 0.5 m between successive passes. Allowing a DVL navigation error of 1% DT, to guarantee sufficient overlap while making sufficient progress, this would restrict the width of the survey area (i.e. the length of successive lawnmower swaths) to only 25 m.

Thus, in order to take advantage of vision as a sensor, it must be possible to register images not only to the most recently snapped reference tile, but also to previously visited tiles. By using the relative displacement between images taken at points distant in time, it is then possible to knock out the drift in odometry measurements, whether from dead-reckoning or vision. The image registration system described above is able to find such links quickly. Figure 2 shows the vehicle trajectory and side-to-side links produced by the vision system for a on online mosaic. The visual side-to-side links, combined with the fast position optimization system described in Section IV, allow the online mosaic and the vehicle position measurement to be corrected and updated on the fly. Thus, the real-time mosaicking and navigation system can produce large mosaics, guaranteeing overlap without resorting to costly external positioning systems.

IV. System Overview

The visual mosaicking and navigation system is divided into two major subsystems: (1) a high-update-rate state measurement and control subsystem, and (2) a loop closure, global error-bounding subsystem which runs asynchronously at a lower, variable rate. The high-update rate system uses the combined vision/DVL odometry system described above and in [10] to determine the current vehicle state for control and to decide when the vehicle has moved far enough that new mosaic tiles should be snapped. The low-update rate loop-closure system takes the new tiles, adds them to the mosaic and attempts to register them with nearby tiles seen previously. Thus it provides side-to-side registration to eliminate drift as described in Section III.

In order to ensure registration between images taken at widely separated points in time, it was found that all tiles in the mosaic should be scaled and potentially rotated to the same viewpoint reference. This reference is defined by the altitude and heading of the first tile of the mosaic. The image registration component requires that the viewpoint of the images be within 10% in altitude and 2° in rotation. To maintain these limits for the visual odometry system, the vehicle is controlled to maintain constant heading and only slowly varying altitude. While this ensures registration is possible image-to-image along track, the limits may be violated for images separated widely in time if altitude has changed or heading-hold is unsteady. Pitching and rolling produce second-order effects and are not compensated for.

The loop closure system uses a variant of the constrained-pose information filter formulation first proposed by Lu and Milios [13]. The position of each tile is taken as a variable to be estimated. The concatenation x of these positions forms the state vector of the mosaic. Measurements of displacement between tiles from the vision registration or from the DVL form constraints, referred to as links, between the individual tile positions, with the strength of each link given by the variance of the measurement (see Table IV). The loop closure system continues the assumptions made by the vision registration that only two components of the vehicle motion—the translation t parallel to the bottom plane—need to be computed. The remaining components of the 6-DOF vehicle state take their

TABLE I

Mosaic link types.

Link Type	*Link Variance*
Vision	From correlation confidence, altimeter variance
DVL	Based on distance traveled (2-σ = 2% DT)
Manual	2-σ = 2 pixels

values from the absolute (drift-free) measurements given by the altimeter, inclinometers, and compass. This greatly reduces the size of the state vector and speeds up computation. In addition, removing the orientations from the state vector makes the system linear, obviating the need for linearizations and/or iterative methods to estimate the state.

The information filter formulation brings additional computational benefits and flexibility. The information filter is the dual of the Kalman filter. Where the Kalman filter estimates the state and covariance (x, P) of a system directly, the information filter works with their duals (ξ, Ω) where

$$\Omega = P^{-1}, \qquad \xi = \Omega x.$$

For the Kalman filter, the state update is fast, whereas a measurement update involves inverting the estimated covariance matrix $\hat{P}$, which makes it impractical for systems with frequent measurements and large state vectors. On the other hand, for the information filter, the state update involves inverting $\hat{\Omega}$, while the measurement update is fast. Thus the information filter is particularly suited to the mosaic loop closure problem, as the locations of the image tiles are assumed to be constants, so that there is no state update, while measurements, in the form of new links between image tiles, can be quickly incorporated. The measurement update involves incrementing the elements of $\hat{\xi}$ and $\hat{\Omega}$ corresponding to the two tiles being linked, thus adding information about those states to the estimate. As a consequence, $\hat{\Omega}$ is sparse: the only non-zero elements are along the diagonal and at locations corresponding to linked tiles.

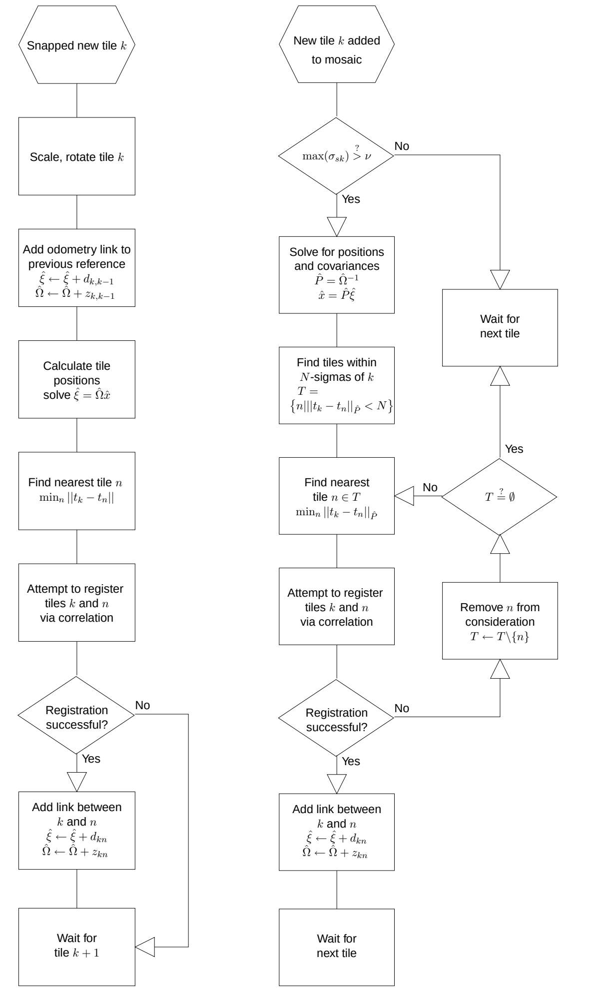
The information filter does not give an immediate estimate of the tile positions. However, the tile positions $\hat{x}$, can be quickly recovered by solving the sparse symmetric positive definite system

$$\hat{\xi} = \hat{\Omega}\hat{x}$$

for $\hat{x}$. Solving this system at any point in time returns an optimal estimate of the positions of all tiles given all the inter-tile links accumulated up until that point.

When the visual odometry system snaps a new tile, with index k, it is added to the mosaic and an attempt is made to find a side-to-side link. This procedure is shown in Figure 3(a), and proceeds as follows :

1) Using altimeter and compass measurements, the tile is scaled and rotated to the reference viewpoint.
2) The displacement information $d_{k,k-1}$ and measurement inverse covariance $z_{k,k-1}$ between tile k and its predecessor are added as a measurement update to the information filter, and the scaled and rotated image is added to the list of mosaic tiles.



(a) Side-to-side linking.

(b) Fallback search.

Fig. 3. Loop closure link-finding procedures, as explained in Section IV.

3) The system is solved for an updated estimate of tile positions, also incorporating information from any previous side-to-side links.
4) Based on the current estimate, the tile n closest to tile k is selected.
5) Tiles k and n are passed to the image registration routine.
6) If the images registered successfully, the information filter is updated with th resulting measurement (d_{kn}, z_{kn}), otherwise, no update is performed, and the system waits for the next tile.

It is possible that many successive side-to-side link attempts fail. This can be due to a lack of side-to-side overlap between tiles, excessive relief in the area of overlap making the two views highly dissimilar, or to occluding fauna, dust or lack of texture in the overlapping region. In this case, the odometry measurement may drift enough that the closest tile to the current tile position estimate is not the correct registration candidate. To overcome measurement drift during periods of extended side-to-side link failure, another slower but more thorough fallback search routine runs periodically after a tile has been added to the mosaic as described above. The variance σ_{sk}^2 in the measurement of the position of the current tile k since the last successful side-to-side link at tile s—i.e. the estimated drift in the odometry—is tracked. If the maximum component of this variance exceeds a threshold ν, it is assumed that the quick search is getting lost and the fallback search is triggered. This algorithm is shown in Figure 3(b) and is as follows :

1) Check whether the position variance exceeds ν.
2) If not, do nothing. If so, solve for the estimated position $\hat{x}$ and its covariance $\hat{P}$. This is an expensive operation as it requires fully inverting the information matrix $\hat{\Omega}$.
3) Find a list T of tiles which fall within a Mahalanobis distance threshold N of the current tile.
4) Step through this list in the order of increasing Mahalanobis distance and attempt to register successive tiles.
5) If a successful registration is found, add the resulting link to the information filter and stop the search.

If this search is unable to find links to knock out the odometry drift, a warning is issued to the user, who can ignore it, and let the fallback search re-run at a later time, or can aid the system by manually matching tiles.

V. Results

The system described here has been fielded on the MBARI ROVs *Ventana* and *Tiburon* to provide a real-time view of the seafloor environment, and to control the vehicle in that environment. These online navigation-grade mosaics have been created at various sites.

Figure 1 shows a small-scale mosaic of an area of scientific interest. It was found that it is not necessary to perform the side-to-side link search for every tile, as can be seen in Figure 2. Though the automatic image registration failed for certain portions of the run, this was due to large parallax resulting from the skeleton relief, or to lack of seafloor visibility.

Figure 4 shows a much larger field which was successfully mosaicked with the system. The mosaic covers approximately 1000 m^2 with 2136 tiles. The only operator input required to form the mosaic was to issue the necessary vehicle position commands.

VI. Conclusion

The real-time mosaicking and navigation system described in this paper can eliminate the need for an external LBL system without suffering from the time-unbounded drift exhibited by dead-reckoning sensors, and, additionally, gives an immediate navigation-grade mosaic providing a view of areas the vehicle has visited.

VII. Acknowledgment

The authors would like to thank the Monterey Bay Aquarium Research Institute for their support in performing this research, including seatime with both the *Ventana* and *Tiburon* ROVs.

Fig. 4. Portion of a real-time, navigation-grade mosaic of a brachiopod/rubble field in Monterey Bay taken from ROV *Ventana*, controlled by the real-time mosaicking and navigation system. The camera altitude from bottom was held at 1.8 m. The total area covered was about 13 m $\times$ 72 m, containing 2136 tiles. The mosaic was created without operator input beyond issuing position commands to move the vehicle. This view shows the top half of the surveyed field, covering an area of about 13 m $\times$ 37 m with 1100 tiles.

Fig. 5. Inset view of the portion of Figure 4 highlighted in yellow. The view measures 7 m $\times$ 7 m.

References

[1] H. Singh, J. Howland, D. Yoerger, and L. Whitcomb, "Quantitative photomosaicking of underwater imagery," in *Oceans 1998*, vol. 1. IEEE, Sept. 29–Oct. 31 1998, pp. 263–266.

[2] S. D. Fleischer, R. L. Marks, S. M. Rock, and M. J. Lee, "Improved Real-Time Video Mosaicking of the Ocean Floor," in *Proceedings of the OCEANS 95 Conference*. San Diego, CA: MTS/IEEE, October 1995, pp. 1935–1944.

[3] R. Marks, S. Rock, and M. Lee, "Real-time video mosaicking of the ocean floor," *IEEE Journal of Oceanic Engineering*, vol. 20, no. 3, pp. 229–241, July 1995.

[4] R. Garcia, J. Puig, P. Ridao, and X. Cufi, "Augmented state Kalman filtering for AUV navigation," in *International Conference on Robotics and Automation, ICRA 2002*. Washington, DC: IEEE, May 2002.

[5] R. Eustice, H. Singh, J. Leonard, M. Walter, and R. Ballard, "Visually navigating the RMS Titanic with SLAM information filters," in *Robotics: Science and Systems (RSS)*, Cambridge, MA, June 2005.

[6] N. R. Gracias, S. van der Zwaan, A. Bernardino, and J. Santos-Victor, "Mosaic-based navigation for autonomous underwater vehicles," *IEEE Journal of Oceanic Engineering*, vol. 28, no. 4, pp. 609–624, Oct. 2003.

[7] S. Negahdaripour and P. Firoozfam, "Positioning and photo-mosaicking with long image sequences; comparison of selected methods," in *Oceans 2001*, vol. 4. MTS/IEEE, Nov. 5–8 2001, pp. 2584–2592.

[8] H. K. Nishihara, "Practical realtime imaging stereo matcher," *Optical Engineering*, vol. 23, no. 5, pp. 536–545, Oct. 1984, also in *Readings in Computer Vision: issues, problems, principles, and paradigms*.

[9] S. D. Fleischer, "Bounded-Error Vision-Based Navigation of Autonomous Underwater Vehicles," Ph.D. dissertation, Stanford University, Stanford, CA 94305, May 2000.

[10] K. Richmond and S. Rock, "A real-time visual mosaicking and navigation system," in *Proceedings of the Unmanned Untethered Submersible Technology Conference (UUST)*. Durham, NH: AUSI, Aug 2005.

[11] L. Whitcomb, D. Yoerger, and H. Singh, "Advances in doppler-based navigation of underwater robotic vehicles," in *International Conference on Robotics and Automation, ICRA 1999*, vol. 1. Detroit, Michigan: IEEE, May 1999, pp. 399–406.

[12] R. McEwen, H. Thomas, D. Weber, and F. Psota, "Performance of an AUV navigation system at arctic latitudes," in *Oceans 2003*. MTS/IEEE, Sept. 22–27 2003.

[13] F. Lu and E. Milios, "Globally consistent range scan alignment for environment mapping," *Autonomous Robots*, vol. 4, pp. 333–349, Apr. 1997.