1998 IEEE International Conference on Intelligent Vehicles, October 1998

Experimental Validation of a Real-Time Vision Sensor and Navigation System for Intelligent Underwater Vehicles

Stephen D. Fleischer *† Stephen M. Rock ‡§
Stanford University Aerospace Robotics Laboratory
Durand Building 250
Stanford, California 94305

Abstract

This paper will present recent advances in the development of a vision sensor for automatic control of underwater robots. This research is a joint effort between the Aerospace Robotics Laboratory (ARL) at Stanford University and the Monterey Bay Aquarium Research Institute (MBARI). Specifically, we have developed and verified experimentally a vision sensor which provides real-time vehicle global position updates, while simultaneously creating a mosaic, or composite image map, of the ocean floor. This novel sensor is the enabling technology within our vision-based navigation system for intelligent underwater vehicles.

I. INTRODUCTION

Our primary motivation for the development of a vision-based sensor for autonomous navigation from video has been to enhance the capabilities of underwater robots used by marine scientists. With this goal in mind, we have successfully demonstrated in our previous research the use of vision as a local displacement sensor in unstructured, underwater environments. Specifically, we have accomplished two crucial tasks using our vision sensor: station-keeping [9] (i.e. holding station over a fixed point on the ocean floor), and mosaicking [8] (i.e. forming a composite image of the ocean floor by aligning successive images from a video stream). We have demonstrated these tasks in both the test tank with *OTTER*, an autonomous underwater vehicle (AUV) designed and built by students at ARL and technicians at MBARI, and in the open ocean on *Ventana*, a remotely operated vehicle (ROV) owned by MBARI.

The focus of our current work is to extend our vision capabilities from local displacement to global position sensing. (In this context, "global" refers to measurements taken with respect to a reference frame which is fixed to the ocean floor, and whose origin is defined by the initial location of the vehicle. In contrast, a "local" measurement refers to the relative displacement between a pair of images.) While our advances in video mosaicking have been useful for both AUVs and ROVs, this task suffers from a serious limitation, namely, the propagation of image alignment errors as the length of the mosaic increases (see Figure 1). This error propagation has two detrimental effects :

- It restricts the use of vision as a global position sensor, since dead-reckoning (i.e. the integration of local image displacement measurements to determine global position) will result in unbounded errors in the global measurements over time.
- It places a practical limit on the size and accuracy of the final mosaic, since the errors will result in poor image alignment whenever the vehicle path crosses back upon itself. This mosaic is the map used for both vehicle control and user specification of navigation commands.

To overcome these obstacles, we have developed a method for optimal estimation of the vehicle's global position. The key sources of information in this technique are the crossover points in the image chain, where the vehicle path loops back upon itself (Figure 1). Our method takes advantage of these crossover points to improve the position estimates in two stages:

- **Crossover Detection and Correlation** By detecting when the vehicle path crosses itself, and correlating the two images at each

---

*Doctoral Candidate, Department of Aeronautics and Astronautics, fleisch@sun-valley.Stanford.EDU

†Research Assistant, Monterey Bay Aquarium Research Institute

‡Associate Professor, Department of Aeronautics and Astronautics, rock@sun-valley.Stanford.EDU

§Adjunct Senior Research Engineer, Monterey Bay Aquarium Research Institute

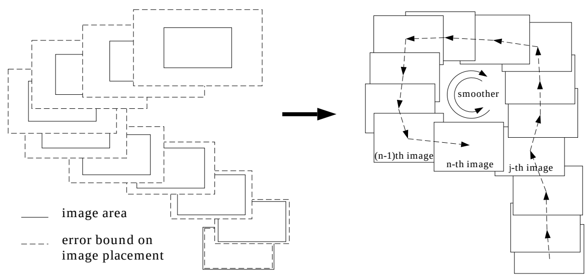

Figure 1: **Error Propagation, Reduction in Image Chain**

For an image chain of length n, the error variance in the global position of the final image is proportional to n. By utilizing optimal estimation, the additional information gained at the crossover point can be propagated along the entire loop to minimize the errors along that section of the image chain.

crossover point, we can obtain an additional measurement of the vehicle's current global position. This external measurement can be used to "reset" the dead-reckoning integration error, thereby improving future position estimates (Figure 2).

- **Smoother** By utilizing the additional estimate of current global position (from the previous stage) as a starting point, we can propagate our sensor model equations backwards to obtain a second estimate of the global position and error at every point along the loop. By combining these estimates with the forward estimates performed in real-time, we can reduce the position errors around the entire loop (Figure 2), thereby improving the accuracy of the mosaic map.

Recently, we have verified experimentally our optimal estimation method on our precision gantry platform, a system capable of controlling a camera head in 6 DOF within its workspace. The development and experimental validation of these core algorithms will be detailed in this paper.

The utility of this new vision technology extends beyond the arena of underwater vehicle control - this work can also be applied to land, air, and space-based vehicle control and navigation applications.

II. BACKGROUND

In just the past few years, the use of vision as a sensor for control and navigation has been a topic of research on a wide variety of vehicle platforms. These range from video correlation-based terrain-tracking techniques for planetary lander navigation [10], to visual landmark navigation for land-based vehicles

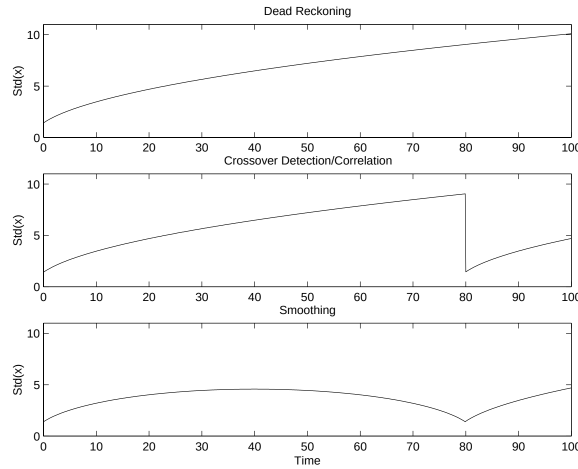

Figure 2: **Stages of Position Estimation**

Stage 1: Sequence the kinematic state equations forward to obtain dead-reckoned position estimates. Stage 2: Detect crossover points in the vehicle path, and correlate the overlapping images to "reset" the dead-reckoned integration error. Stage 3: Smooth the estimates around the loop in the image chain, by propagating the state equations backward and combining these measurements with the forward estimates.

[13]. The latter work is akin to our current research in that it uses optimal estimation techniques, in this case an extended Kalman filter (EKF), to improve vehicle position estimates for the purpose of autonomous navigation. Their particular focus is the maximum likelihood detection and identification of visual features (i.e. landmarks) using current vehicle state estimates. The resulting measurements are then applied to future state estimates through the EKF. In essence, their vision measurements are used to improve the dead-reckoned state estimates derived from on-board sensors, which is analogous to the role of crossover detection and correlation in our work.

Recently, several researchers have approached the problem of underwater mosaicking, to produce visual maps of the ocean floor. Interesting results have been achieved in the area of constrained video mosaicking, in which a multiple-column mosaic is created by correlating the images in adjacent columns [8]. This research effort has also produced impressive single-column mosaics of the sea floor using *Ventana*, the MBARI ROV. Related research has investigated the possibility of extending the concept of mosaicking to include 3-D motion estimation for vehicle control [12]. There has also been promising theoretical work to solve the occlusion problem when mosaicking terrain with significant altitude variations [6, 11]. While this research has been quite successful,the intensive computations required prohibit a real-time implementation of the necessary algorithms.

Our current work builds upon the latest successes in video mosaicking, in an attempt to produce a system capable of creating maps not just for visualization, but for the purposes of vehicle position sensing and navigation. While recent simulations have been quite successful in demonstrating terrain-relative underwater navigation by matching current video or sonar data to bathymetry maps of the ocean floor [2], our goal is to explore unknown environments, without the benefit of any *a priori* knowledge of the scene. Our past work has explored several alternate optimal estimation techniques [3, 5, 4], including both discrete and continuous dynamic filter-smoothers. However, these algorithms were limited in their practical usefulness, either because they relied on restrictive assumptions, such as constant inter-image local displacements or the presence of a dynamic model of the vehicle, or because they were too computationally intensive to be implemented in an on-line navigation system. This paper describes several algorithms which combine the strengths of our previous approaches to enable on-line corrections to the mosaic and current vehicle state estimate, based on the vision sensor kinematics. Dynamic models of the vehicle can be used to further improve the estimates, but they are not required by our current method.

III. OPTIMAL ESTIMATION METHOD

This section will provide an overview of our method for optimally estimating the vehicle global position from image data. Before proceeding, it is important to understand the mosaicking process. At initialization, a snapshot of the current ocean floor scene is stored as a static reference image. Every new image from the camera is then correlated against this image by comparing texture regions in both images, to provide a direct measurement of the local image displacement. Whenever the vehicle moves more than one image frame, a new reference image is snapped and aligned with the previous one, thereby creating a composite image of the scene as the vehicle moves. Currently, vehicle motion is restricted to 2-D translational motions within a plane parallel to the ocean floor.

A. Sensor Model

Once these local displacements have been measured, they are used as inputs to the sensor state model to estimate image and vehicle global displacements. This sensor model is a set of kinematic equations based on the sensor geometry. Figure 3 depicts this geometry, where W is the world frame, C and V correspond to the image and vehicle frames, respectively, when the reference snapshot was taken, and C' and V' correspond to the current image and vehicle frames, respectively. I and I' (not shown) are the 2-D frames in which the inter-image local displacements are measured. Thus, the I, C, and V and frames are updated every time a new snapshot is taken, and the I', C', and V' frames are updated at approximately 15 Hz. Units are pixels in the I frames; units are meters/radians in the C and V frames. Based on this geometry, a first order approximation of the image and vehicle global state and error covariance has been derived, using the following assumptions:

- The local displacements ${}^I\delta x$, ${}^I\delta y$, the range ${}^{C'}r$, and the Euler angles ${}^W\phi_{C'}$ (roll), ${}^W\theta_{C'}$ (pitch), and ${}^W\psi_{C'}$ (yaw) are the available, independent measurements at each time step, along with their corresponding variances. Furthermore, all measurements at time k are independent from those at time $k+1$.

- The error distributions on these fundamental measurements can be modeled as Gaussian.

- The world (i.e. global) frame is defined such that its origin is coincident with the origin of the initial image, and its axes coincide with the vehicle frame axes when ${}^V\phi = {}^V\theta = {}^V\psi = 0$.

- The ocean floor is flat and planar, and it coincides with the X-Y plane of the world frame.

- The vehicle is free to translate (not rotate) in a single plane parallel to the ocean floor.

- All rotations (${}^W\phi_{C'}$, ${}^W\theta_{C'}$, and ${}^W\psi_{C'}$) are small ($<$ .1 rad).

- $\frac{{}^C\delta x}{{}^C r}$, $\frac{{}^C\delta y}{{}^C r}$ are small ($<$ .1).

- $Var({}^C r) << \mu^{C}\hat{r}^2$, $Var({}^W\psi_{C'}) << 1$ rad^2

Since the equations that result from this derivation are too lengthy to describe in this paper, it is sufficient for this discussion to state that the outputs of this model are the image global state estimate vector (${}^W\hat{X}_{C'}$), the image global state error covariance matrix ($Cov({}^W X_{C'})$), the vehicle global state estimate vector (${}^W\hat{X}_{V'}$), and the vehicle global state error covariance matrix ($Cov({}^W X_{V'})$). The image state vector is defined as follows:

$$
{}^{W}\hat{X}_{C'} \equiv \begin{pmatrix} {}^{W}\hat{x}_{C'} \\ {}^{W}\hat{y}_{C'} \\ {}^{W}\hat{z}_{C'} \equiv 0 \\ {}^{W}\hat{\phi}_{C'} \\ {}^{W}\hat{\theta}_{C'} \\ {}^{W}\hat{\psi}_{C'} \end{pmatrix}
$$

and the vehicle state vector is similarly defined.

This set of kinematic equations is propagated forward as new images are added to the evolving mosaic. Thus, real-time estimates of image and vehicle state are produced as new image measurements are received. As shown in Figure 5, the variance of these dead-reckoned state estimates increase without bound as the path length increases.

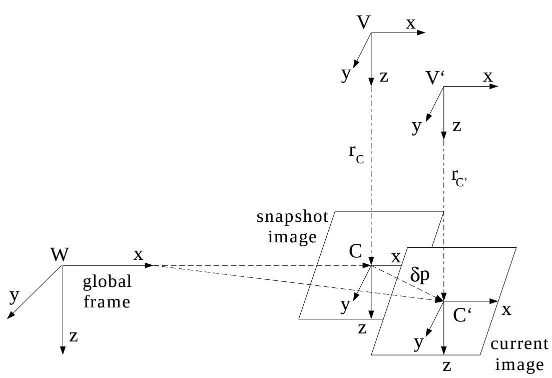

Figure 3: **System Geometry**

B. Crossover Detection and Correlation

In order to determine exactly where the vehicle path intersects itself, the location of every image in the mosaic is checked, relative to the current image. However, due to the uncertainty of the image position estimates, it is necessary to determine the bounding ellipsoid of possible locations for each image at every time step.

The first step in determining these bounds is to propagate the kinematic state equations backwards from the current image through the image chain. Note that since only the location of each image relative to the current image is required, the sensor model is initialized with zero position and variance. The next step is to compute the $n\sigma$ uncertainty ellipsoids around each relative image location, where n sets the desired confidence in the crossover detection, and σ is the standard deviation of the measurement. For instance, if $n = 3$, there is a 98.9% probability that the actual relative image location is within an ellipsoid centered at the estimated location, with semi-axes $3\sigma_x$, $3\sigma_y$. If an image is found such that its location ellipsoid lies completely within the current image, this image pair is considered to be a crossover point.

Once a crossover image-pair is detected, the two images are correlated to measure their local displacement. Using this measurement, a new estimate of the current image global position is calculated, which is used in future position estimation calculations as the vehicle continues along its path.

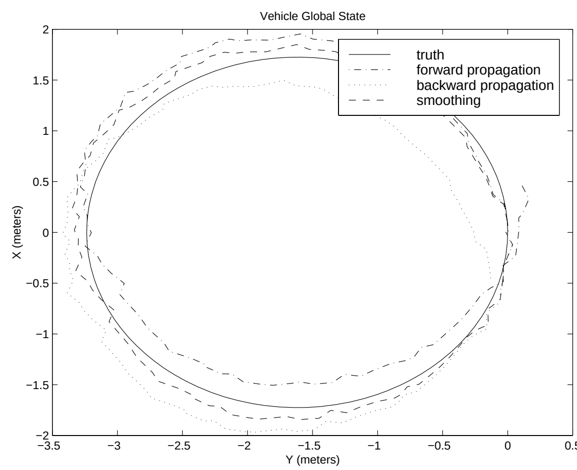

Figure 4: **Vehicle State Estimates**

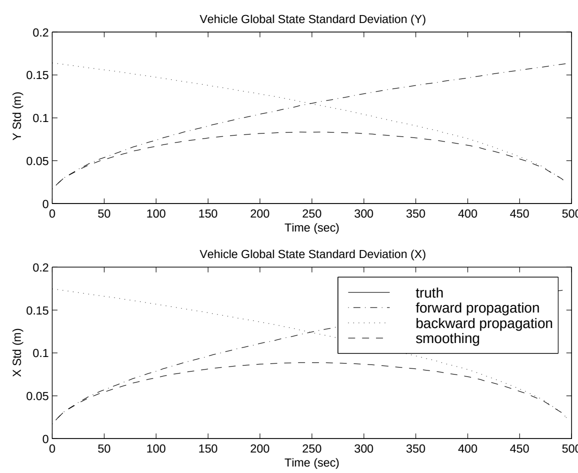

Figure 5: **Vehicle State Variance Estimates**

C. Smoother

Given the additional estimate of the current image position, it is possible to use this measurement to improve the estimates around the loop just detected. First, the kinematic sensor equations are propagated backwards, using the crossover estimate of current image position as a starting point (Figures 4, 5). Thus, for each point along the loop, there are two estimates of position, from forward and backward propagation of the sensor model. Note that these two estimates are independent at every point k, since the forward estimate uses local displacement measurements 1 through k, while the backward estimate uses local displacement measurements n through $k+1$, where $n+1$ is the length of the image chain. At each point k, these two measurements can be combined to determine the maximum likelihood estimate of image global position [1]:

$$S(k) = S_b(k) + S_f(k)$$

$$x(k) = S^{-1}(k)\left[\lambda_b(k) + \lambda_f(k)\right]$$

where the information matrix $S \equiv Cov(x)^{-1}$ and the information vector $\lambda \equiv Sx$. These improvements are demonstrated in Figure 4, where the dead-reckoned and smoothed estimates are compared to the truth, and Figure 5, which shows the reduction in estimate error variance.

V. EXPERIMENTAL RESULTS

A. Experimental Mapping and State Estimation

We have recently demonstrated our mosaicking and vehicle global position estimation capabilities on our 6-DOF precision gantry platform. Using a dual-Pentium 133 MHz PC, we were able to estimate the vehicle global position at 15 Hz, while a mosaic of the scene was created (Figure 6). During this computation, the evolving mosaic was displayed in real-time to the user, including graphical overlays representing the current vehicle position estimate and uncertainty ellipsoid.

B. Crossover Detection Validation

In addition to demonstrating our mosaicking technology, we were able to implement the first of the algorithms described in this paper, namely, crossover detection. During the creation of the mosaic shown in Figure 6, the vision sensor propagated the sensor equations backwards at every time step and checked for overlap with any of the existing images within the mosaic. As the mosaic looped back upon itself, the crossover between the first and final images was detected. As our research progresses, we will continue to implement these core algorithms and test them experimentally, using our gantry as a truth measurement system to verify our results.

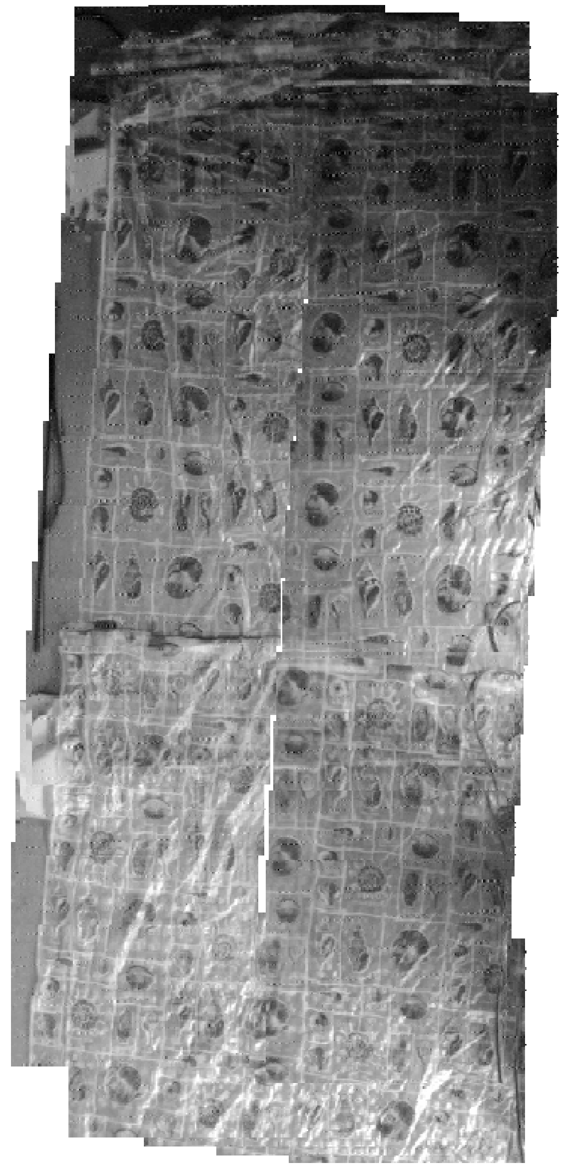

Figure 6: **Experimental Mosaic**

C. Navigation System

In previous work, we have demonstrated experimentally the ability to navigate *OTTER* from vision in the MBARI test tank environment [7]. The next phase of our research will be to incorporate our optimal estimation techniques into our navigation system, and demonstrate this improved technology both on our precision gantry platform, and on *OTTER* in the test tank.

V. CONCLUSIONS

We have presented a novel method for the optimal estimation of vehicle global position from a vision sensor. In particular, two core algorithms were discussed: the crossover detection and correlation, and the smoother. The validity of this approach was demonstrated experimentally for the first time on our 6-DOF precision gantry platform. This technique is the key enabling technology for our future demonstrations of robust, autonomous underwater vehicle navigation in unknown environments.

REFERENCES

[1] Bryson, A. Dynamic Optimization with Uncertainty. Paper in progress, March 1993.

[2] di Massa, D., and Stewart, W. Terrain-Relative Navigation for Autonomous Underwater Vehicles. In *Proceedings of the OCEANS 97 Conference* (1997), MTS/IEEE.

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

[4] Fleischer, S. D., Rock, S. M., and Burton, R. L. Global Position Determination and Vehicle Path Estimation from a Vision Sensor for Real-Time Video Mosaicking and Navigation. In *Proceedings of the OCEANS 97 Conference* (Halifax, Nova Scotia, October 1997), MTS/IEEE.

[5] Fleischer, S. D., Wang, H. H., Rock, S. M., and Lee, M. J. Video Mosaicking Along Arbitrary Vehicle Paths. In *Proceedings of the Symposium on Autonomous Underwater Vehicle Technology* (Monterey, CA, June 1996), OES/IEEE, pp. 293–299.

[6] Hert, S., Tiwari, S., and Lumelsky, V. A Terrain-Covering Algorithm for an AUV. *Autonomous Robots 3*, 2-3 (June-July 1996), 91–119.

[7] Huster, A., Fleischer, S. D., and Rock, S. M. Demonstration of a vision-based dead-reckoning system for navigation of an underwater vehicle. In *Proceedings of the OCEANS 98 Conference* (Nice, France, September 1998), IEEE. submitted.

[8] Marks, R., Rock, S., and Lee, M. Real-time video mosaicking of the ocean floor. *IEEE Journal of Oceanic Engineering 20*, 3 (July 1995), 229–241.

[9] Marks, R. L., Wang, H. H., Lee, M. J., and Rock, S. M. Automatic visual station keeping of an underwater robot. In *Proceedings of IEEE Oceans 94 Osates* (Brest, France, September 1994), IEEE.

[10] Reid, M. B., and Hine, B. P. Terrain Tracking for Lander Guidance Using Binary Phase-Only Spatial Filters. In *Photogrammetric Engineering and Remote Sensing* (December 1992).

[11] Tiwari, S. Mosaicking of the Ocean Floor in the Presence of Three-Dimensional Occlusions in Visual and Side-Scan Sonar Images. In *Proceedings of the Symposium on Autonomous Underwater Vehicle Technology* (June 1996), OES/IEEE.

[12] Xu, X., and Negahdaripour, S. Vision-based Motion Sensing for Underwater Navigation and Mosaicing of Ocean Floor Images. In *Proceedings of the OCEANS 97 Conference* (1997), MTS/IEEE.

[13] Yoder, J.-D., Skaar, S., and Arriola, H. Using Probability Estimates to Identify Environmental Features for a Nonholonomic Control System. *Journal of Guidance, Control, and Dynamics 20*, 6 (Nov.-Dec. 1997), 1215–1220.