Global Position Determination and Vehicle Path Estimation from a Vision Sensor for Real-Time Video Mosaicking and Navigation

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

Richard Burton [‡]
Monterey Bay Aquarium Research Institute
P.O. Box 628
7700 Sandholdt Road
Moss Landing, California 95039-0628

Abstract

This paper will describe our continuing advances in a joint research effort, undertaken by the Aerospace Robotics Laboratory (ARL) at Stanford University and the Monterey Bay Aquarium Research Institute (MBARI), to enable autonomous navigation from video for unmanned underwater vehicles. In particular, we have developed a real-time vision system for vehicle position sensing and sea-floor mapping. We will present the theoretical development of this work and associated simulation results. In addition, we will discuss the experimental verification of our techniques in the lab environment.

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 [6] (i.e. holding station over a fixed point on the ocean floor), and mosaicking [5] (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.

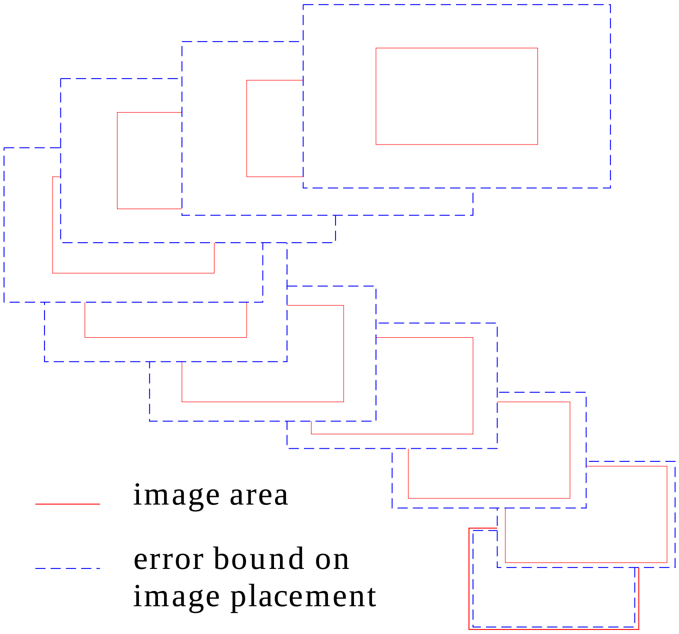

Figure 1: **Error Propagation in Image Chain**

For an image chain of length n, the error variance in the global position of the final image is proportional to n.

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" measure-

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

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

[‡] Senior Research Engineer, Monterey Bay Aquarium Research Institute, budi@Mbari.ORG

ment 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 Fig. 1). This precludes 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. Furthermore, this error propagation places a practical limit on the size of the final composite image, since the errors will result in poor image alignment whenever the vehicle path crosses back upon itself. While we have partially solved this problem by using a discrete method which reduced errors around loops in the vehicle path [2, 3], this algorithm placed several artificial constraints on the vehicle and vision processing subsystem. These constraints will be discussed in detail in this paper.

To overcome these obstacles, we have developed a continuous method for error reduction, which computes both the global vehicle position at snapshot times and the estimated vehicle path between successive snapshot images. We will present a theoretical development of this algorithm and associated simulation results in this paper. We will also demonstrate its capabilities with the aid of a simple pan/tilt camera setup in the lab environment. Our vision technology has been developed to be a “black box” sensor. As such, its utility 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

Over the past decade, several teams of researchers have made significant advances in utilizing various computer vision and image processing techniques for sensing and control of underwater vehicles. Our current research is the latest in a progression of experiments towards better underwater vehicle control through vision sensing.

One of the first algorithms to be tested in the marine environment was the computation of optical flow. With this method, vehicle motion is determined by comparing features, lines, or textures between the current and previous images in a video stream. This pseudo-velocity measurement has been used in the past to control an underwater vehicle, specifically for the task of station-keeping [7, 8, 11]. However, in order to achieve position control, this velocity sensor had to be integrated, thereby resulting in dead-reckoning error which grew rapidly as a function of time (frame rates of 15-30 Hz are typical).

To achieve the task of drift-free station-keeping, refinements to the vision sensor were accomplished by modifying the optical flow measurement technique. Instead of correlating the latest two images in the video stream, a snapshot of the desired hover point is stored as a static reference image. Every new image from the camera is then correlated against this image, to provide a direct measurement of vehicle displacement. While this approach proved to be quite successful [4, 6, 10], the magnitudes of vehicle motions are limited to a single image frame. As a result, large disturbances result in a loss of station, and no measurement of vehicle position, local or global, is available outside the reference image.

Recently, several researchers have approached the problem of underwater mosaicking, to produce visual maps of the ocean floor. By extending the station-keeping method to snap a new reference image whenever the vehicle moves more than one image frame, and align this new snapshot image with the previous one, a composite image of the scene is created as the vehicle moves. 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 [5]. This research effort has also produced impressive single-column mosaics of the sea floor using *Ventana*, the MBARI ROV. There has also been promising theoretical work to solve the occlusion problem when mosaicking terrain with significant altitude variations [9]. 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.

III. ERROR REDUCTION METHOD

In order to adapt our video mosaicking task to produce vehicle global position estimates with small, finite errors for any vehicle motion (i.e. along arbitrarily long vehicle paths), it is necessary to reduce the image alignment errors which propagate through the image chain (Fig. 1). To achieve this, we need an external measurement of global position, in order to “reset” the dead-reckoning integration error (Fig. 2). This can be considered to be a sensor fusion problem for two dissimilar measurements, with two important distinctions. First of all, it is not sufficient to simply update the current global position and continue the dead-reckoning process. Because the final mosaic will be used as a correlation reference map for vehicle navigation, this improvement must be propagated back through the image chain, to improve the global placement of all images within the mosaic. This will also have the beneficial side effect of improving the visual quality of the mosaic. Second, we can obtain an external measurement of the vehicle global position whenever the vehicle path crosses itself, by correlating the two images at the crossover point. Thus, the entire process can be completely encapsulated within the vision sensor.

In our previous work [2, 3], we have developed a dis-

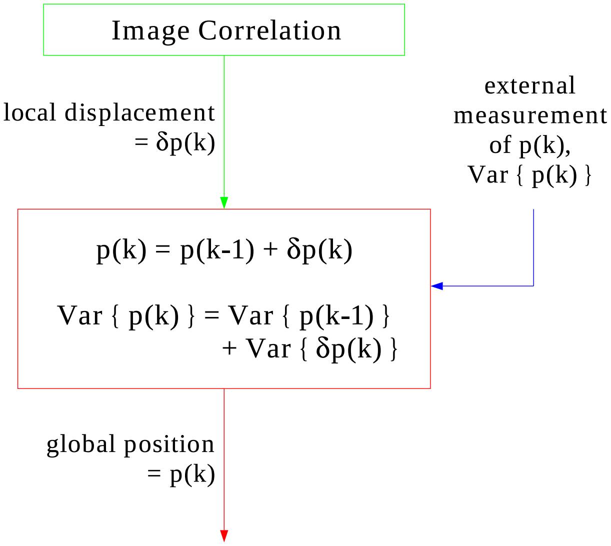

Figure 2: **External Correction to Dead-Reckoning Error**

In order to correct for the integration of errors due to dead-reckoning, an external measurement of global position can be used to "reset" the error variance.

crete method to reduce the errors around loops in the vehicle path (Fig. 3), but there are several limitations to this method, which will be discussed in the next section. The following sections will describe a new continuous method for error reduction. While this method utilizes the same theory of optimal estimation and filter-smoother techniques (in continuous rather than discrete form), it does not suffer from the same restrictions. As a result, this algorithm is more practical for real-world applications, as we have verified experimentally in the lab environment.

A. Limitations of Discrete Method

To appreciate the advantages of our continuous error reduction method, it is useful to briefly describe the limitations of our previous algorithm. The most significant constraint imposed by the discrete implementation was that the magnitudes of local displacements between consecutive images were required to be constant. Since the total path length was used as the independent variable in our algorithm, variations in the image separations violated the assumption that the problem could be discretized. Since it is often beneficial to take a snapshot before the vehicle has moved the required displacement, in order to improve the robustness of the vision measurement, this assumption is often violated severely in practical situations.

A second limitation encountered was the fact that estimates of the vehicle position only existed for times when a new image snapshot was taken. Thus, although image correlation measurements were being produced at 15 Hz, this data was not being utilized during the smoothing computations to reduce errors around loops in the path. While this is not a critical constraint, it does reduce the accuracy of the final global position estimates.

Finally, it proved to be difficult to implement the discrete algorithm around multiple loops in the vehicle path. While the mathematical difficulties were not impossible to overcome, the derivation and implementation is much simpler for the continuous method.

B. Continuous Optimal Estimation Theory

Although an explanation of continuous finite-time filter-smoother theory is beyond the scope of this paper, it is important to recognize how we have applied these techniques to our problem, and how we have adapted the standard equations [1] for our specific geometry.

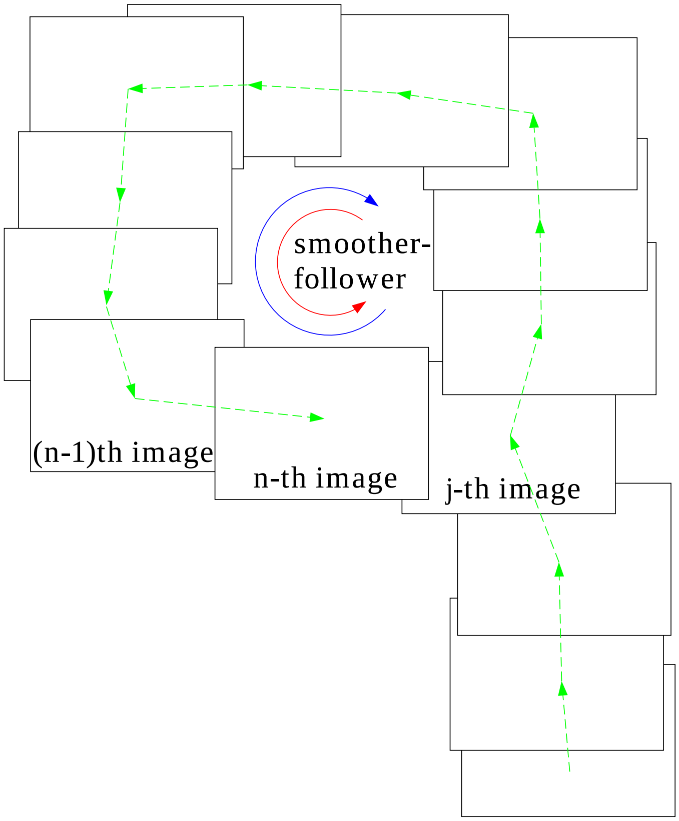

Figure 3: **Geometry for a Closed Loop Image Chain**

By utilizing a filter-smoother, 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.

For our problem, we have derived a dynamic sensor model based on the geometry of the environment and vehicle, but we have no dynamic model of the vehicle plant itself. As a result, it is impossible for Kalman filtering to improve the local or global image position estimates within a mosaic [1]. However, we can gain additional information whenever the image chain loops back upon itself. By correlating the *nth* image with the *jth* image as well as the $(n-1)th$ image (Fig. 3), we gain another measurement of the nth image global state. Furthermore, this new measurement is more accurate, since the jth image occurs earlier in the image chain and thus its global state measurement has a lower variance. By isolating the measurements along the loop between image j and image n, we can propagate this additional information by applying an optimal filter-smoother to these measurements.

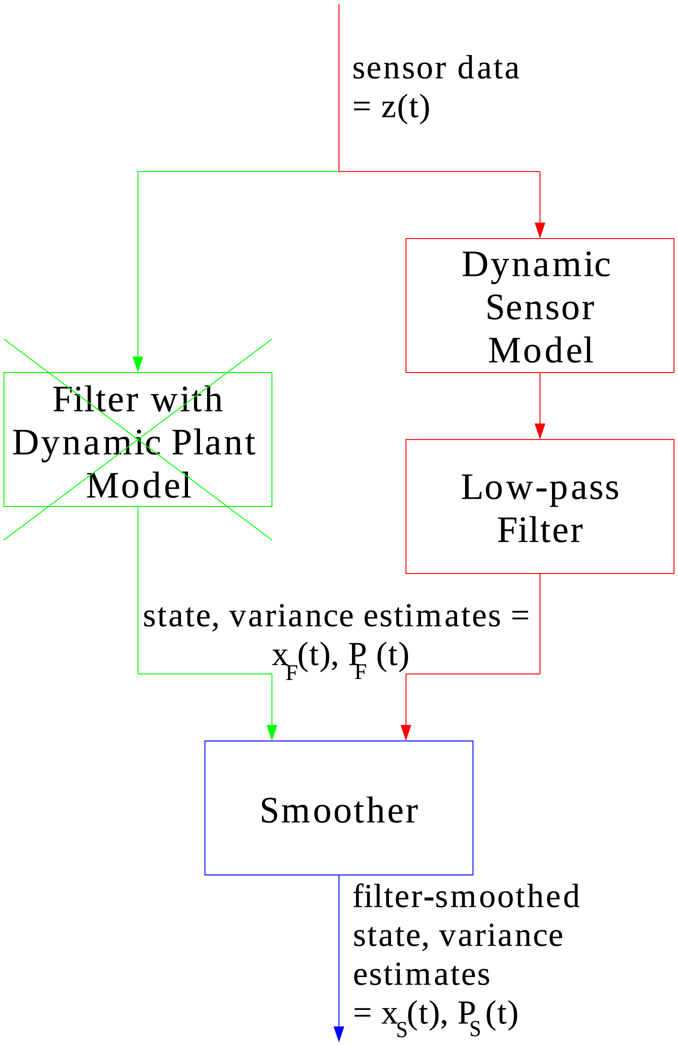
sensor data
= z(t)
Dynamic Sensor Model
Filter with Dynamic Plant Model
Low-pass Filter
state, variance estimates =
x_F(t), P_F(t)
Smoother
filter-smoothed state, variance estimates
= x_S(t), P_S(t)

Figure 4: **Block Diagram of the Continuous Filter-Smoother**

This diagram illustrates how the standard filter, based on a dynamic model of the plant, has been replaced with a dynamic sensor model and a low-pass filter.

Since we do not have a dynamic model for the underwater vehicle, this causes problems in implementing the standard continuous filter-smoother equations. Fig. 4 illustrates the logical structure of the standard algorithm. The filtering phase, which produces real-time, causal estimates of state ($x_F(t)$) and variance ($P_F(t)$), relies on the existence of a dynamic plant model. To solve this problem, we have replaced this set of equations with our own dynamic sensor model. This model, which will be explained in the next section, produces the same state and variance estimates, which are then used by the smoother phase of the algorithm.

C. Sensor Model

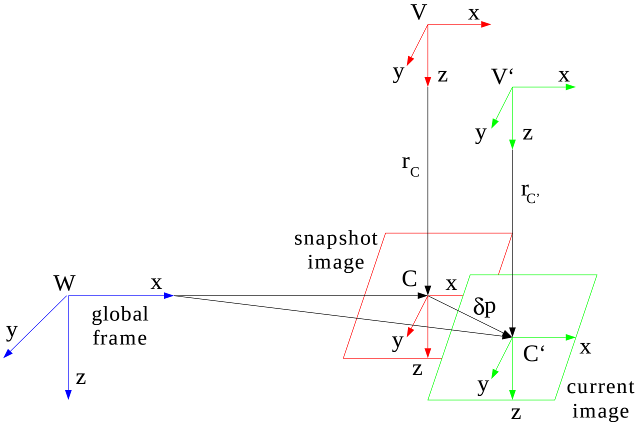

Figure 5: **System Geometry**

The sensor model for our problem can be derived from the system geometry (Fig. 5). In this diagram, the C frame represents the frame of the last snapshot image taken, and the C' frame represents the current image frame. Thus, the C frame is updated every time a new snapshot is taken, and the C' frame is updated at approximately 15 Hz. Based on this geometry, the following iterative dynamic equation can be derived to solve for the position vector ${}^{W}P_{C'}(t)$:

$$\begin{aligned} {}^{W}P_{C'}(t) &= {}^{W}P_{C}(t_{snapshot}) \\ &+ {}^{W}_{C}T(t_{snapshot}) \begin{pmatrix} {}^{C}\delta x_{(m)}(t) \\ {}^{C}\delta y_{(m)}(t) \\ 0 \end{pmatrix} \end{aligned} \qquad (1)$$

where ${}^{C}\delta x_{(m)}$ and ${}^{C}\delta y_{(m)}$ can be determined from the pixel correlation measurements and the camera geometry:

$${}^{C}\delta x_{(m)} = \left[\frac{2r_{C(m)}}{h_{(pix)}} tan\left(\frac{FOV_y}{2}\right)\right] {}^{I}\delta y_{(pix)} \qquad (2)$$

$${}^{C}\delta y_{(m)} = \left[\frac{2r_{C(m)}}{w_{(pix)}} tan\left(\frac{FOV_x}{2}\right)\right] {}^{I}\delta x_{(pix)} \qquad (3)$$

Note the change in coordinates when going from the image measurement frame I (in pixels) to the camera frame C (in meters). FOV_x and FOV_y represent the horizontal and vertical camera fields of view, respectively, while $w_{(pix)}$ and $h_{(pix)}$ represent the width and height of the image in pixels (512 x 480 for our case). The rotation matrix ${}^{W}_{C}T(t_{snapshot})$ is computed from external measurements of the pitch, roll, and yaw of the vehicle. Although a range estimate based on stereo correlations has been derived and implemented, a nominal range $r_{C(m)}$ of 1 meter has been assumed for all data presented in this paper.

In similar fashion, a set of iterative equations for state variance propagation were derived, based on (1) - (3):

$$Var({}^W P_{C'}(t)) \;=\; Var({}^W P_C(t_{snapshot})) \; +$$
$$\begin{pmatrix} V_3 & {}^W\Phi_{C'}(V_3 - V_4) & -{}^W\Theta_{C'}V_3 \\ {}^W\Phi_{C'}(V_3 - V_4) & V_4 & {}^W\Psi_{C'}V_4 \\ -{}^W\Theta_{C'}V_3 & {}^W\Psi_{C'}V_4 & 0 \end{pmatrix} \quad (4)$$

where ${}^W\Phi_{C'}$, ${}^W\Theta_{C'}$, and ${}^W\Psi_{C'}$ are the vehicle roll, pitch, and yaw, respectively, and:

$$V_1 \;=\; Var({}^I\delta x_{(pix)}) \;=\; Var({}^I\delta y_{(pix)}) \quad (5)$$

$$V_2 \;=\; Var(r_{C(m)}) \quad (6)$$

$$V_3 \;=\; Var({}^C\delta x_{(m)}) \;=\; \left[\frac{2}{h_{(pix)}} tan\left(\frac{FOV_y}{2}\right)\right]^2 \times$$
$$\left[r^2_{C(m)}V_1 + {}^I\delta y^2_{(pix)}V_2 + V_1V_2\right] \quad (7)$$

$$V_4 \;=\; Var({}^C\delta y_{(m)}) \;=\; \left[\frac{2}{w_{(pix)}} tan\left(\frac{FOV_x}{2}\right)\right]^2 \times$$
$$\left[r^2_{C(m)}V_1 + {}^I\delta x^2_{(pix)}V_2 + V_1V_2\right] \quad (8)$$

It is important to note that several assumptions have been made during the derivation of this sensor model: (1) the image surface is a plane, and the location and orientation of this plane defines the global reference frame; (2) since the vehicle is under active control, it is reasonable to assume that all angles are small; (3) all error distributions on the pixel correlation measurements are Gaussian, or can be modeled as such.

IV. SIMULATION WORK

To verify the correctness of our sensor model and the optimal smoother, we have tested the algorithms in MATLAB. The results of our simulation study are shown in Figs. 6, 7, and 8. Using a simulated circular vehicle path, Fig. 6 shows the improvement achieved by smoothing the estimates around the loop. Fig. 7 illustrates more closely the difference in accuracy before and after applying our algorithm. As seen in Fig. 8, our method bounds the error around loops in the vehicle path, allowing vehicle paths of arbitrary length (provided there are a sufficient number of crossover points).

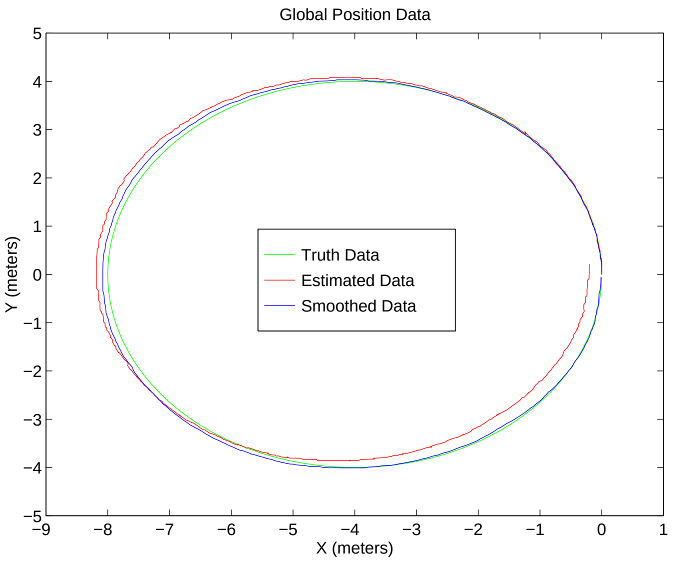

Figure 6: **Global Image Position Along a Circular Vehicle Path**

When compared to the actual image position, the smoothed data is more accurate than the position estimate based purely on noisy sensor data, particularly at the endpoints of the loop.

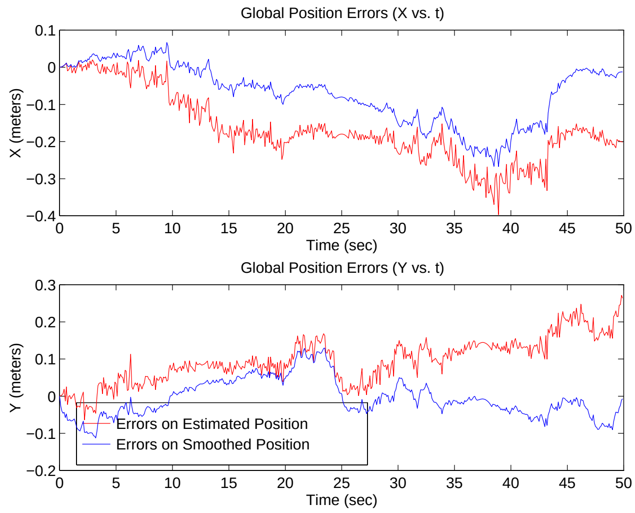

Figure 7: **Image Position Error Along a Circular Vehicle Path**

This plot illustrates more precisely the effect of smoothing the errors around the loop in the vehicle path.

V. EXPERIMENTAL VERIFICATION

We have experimentally demonstrated our algorithm in the lab, using a video camera mounted on a simple pan/tilt tripod. To first order, vehicle translations parallel to the image plane and the corresponding rotations (e.g. side-slip

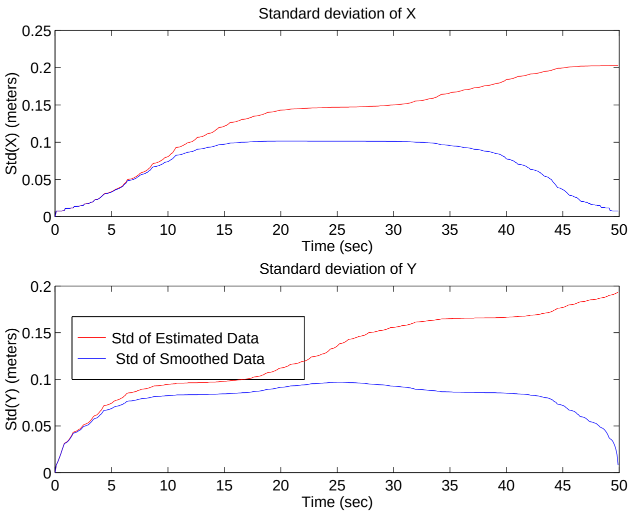

Figure 8: **Standard Deviation of State**

Around any closed loop in the mosaic, our filter-smoother algorithm minimizes the variance, subject to the constraints of the dynamic sensor equations.

and roll) are indistinguishable from each other. Thus, provided that the tripod pan/tilt angles between consecutive image snapshots are small, this system behaves quite similarly to an actual vehicle undergoing planar translations.

The image digitization and real-time correlations are performed by AVP (Advanced Vision Processor), a dual Pentium processor PC which runs proprietary software written by Teleos Research, now a division of Autodesk. The local image displacements are then sent via ethernet to a SUN workstation, which performs global position estimation and smoothing around path loops. This data is then sent back to AVP, in order to update the search parameters for determining crossover points.

A typical mosaic created by this system is illustrated in Fig. 9. After crossover detection, correlation, and smoothing, an improved mosaic is created (Fig. 10). Note the improved image alignment at the crossover point, while good image alignment is maintained throughout the rest of the mosaic. As the size of the mosaic increases (typical areas of the ocean floor to be mapped may be a factor of 10 or 100 greater than the region in this case), the benefits will become even more dramatic.

VI. CONCLUSIONS

We have presented an improved technique for propagating error corrections around loops in image mosaics. By improving the accuracy of image placement within a mosaic, our method enhances the quality of the global position data which can be extracted from the mosaicking process. This tool is an essential component towards the development of a vision-based global position sensor, which will aid in the navigation of underwater robotic vehicles. The next stage of this research will be to obtain quantitative experimental results and determine the performance of our system, and to demonstrate this new capability on an operational underwater vehicle.

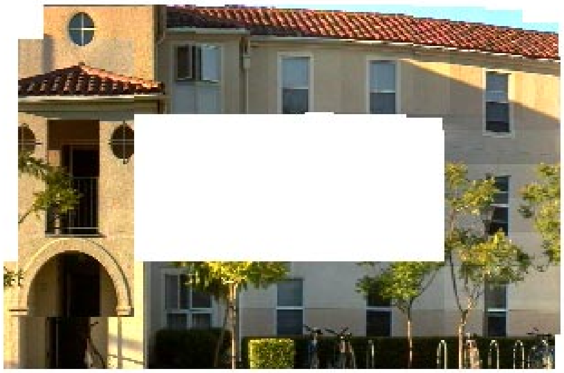

Figure 9: **Unconstrained Mosaic (Estimated Image Alignment Data)**

This mosaic was created autonomously while the camera underwent pan and tilt motions, mimicking an arbitrary vehicle path. Note the poor image alignment between the initial and final images (the archway in the lower left corner).

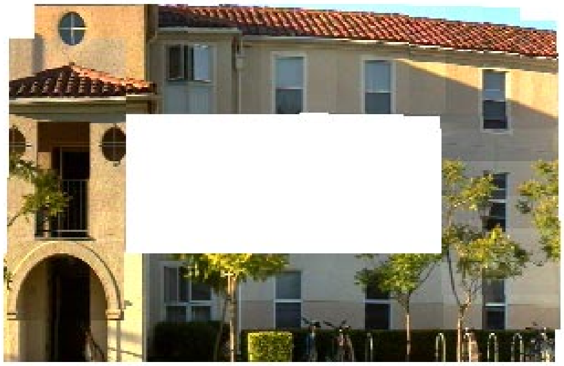

Figure 10: **Unconstrained Mosaic (Smoothed Image Alignment Data)**

This mosaic was created after filtering the image alignment data using our filter-smoother approach. Note the greatly improved alignment between the initial and final images in the mosaic, while the rest of the mosaic has not degraded in quality.

REFERENCES

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

[2] 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.

[3] 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.

[4] Jin, L. Real-Time Vision-Based Stationkeeping System For Underwater Robotics Applications. In *Proceedings of the OCEANS 96 Conference* (1996), MTS/IEEE.

[5] 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.

[6] 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.

[7] Negahdaripour, S. Passive Optical Sensing For Near-Bottom Stationkeeping. In *Proceedings of the OCEANS 90 Conference* (1990), OES/IEEE.

[8] Negahdaripour, S. Undersea Optical Stationkeeping. Improved Methods. In *Journal of Robotic Systems* (June 1991).

[9] 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.

[10] Wasielewski, S. Dynamic Vision for ROV Stabilization. In *Proceedings of the OCEANS 96 Conference* (1996), MTS/IEEE.

[11] Yuh, J. Control and Optical Sensing in Underwater Robotic Vehicles (URVs). In *Proceedings of the OCEANS 90 Conference* (1990), OES/IEEE.