Demonstration of a Vision-Based Dead-Reckoning System for Navigation of an Underwater Vehicle *

Andreas Huster [†] Stephen D. Fleischer [‡] Stephen M. Rock [§¶]
Stanford University Aerospace Robotics Laboratory
Monterey Bay Aquarium Research Institute

Abstract

This paper describes a dead-reckoning navigation system for hover-capable underwater vehicles operating close to the ocean floor. Navigation is presented as an extension of underwater station-keeping and mosaicking. It combines real-time vision-processing to build a mosaic of the area of interest, an image-based user interface to specify desired vehicle locations, and vision-based dead-reckoning to compute the robot's position in the mosaic. This system provides a high-level interface between the vehicle and the pilot, who specifies the goal (e.g., go to and hover over this feature) instead of the commands to execute the task (e.g., rotate to the left, go forward, stop). Thus, it is an enabling technology for autonomous underwater vehicles (AUVs)—for which commanding actuators directly is not feasible—and a useful high-level interface for remotely operated vehicles (ROVs). This new capability is the result of our on-going research with the Monterey Bay Aquarium Research Institute (MBARI).

A. Introduction

This paper presents a vision-based navigation system for hovering underwater vehicles. The focus is on unknown and unstructured underwater environments with few, if any, known landmarks and no computer models. Vision-based navigation is an extension of our previous research in station-keeping and real-time mosaicking.

A scientific exploration mission, where a pilot needs to survey a region of the ocean floor and return to various interesting sites, is an example of where vision-based navigation would be useful. The system generates a mosaic that the pilot can use to point to the location of interest with a mouse or a touch screen. The vehicle also determines its current position by tracking the distance traveled as each new image is added to the mosaic.

A significant feature of navigation-from-video is that the same images can be interpreted by both the vision processor and the pilot, and therefore provide a natural interface between the human operator and the computer.

Figure 1 shows an overview of the navigation system. The vision processor on the vehicle provides two functions: first, it generates and updates a mosaic (a composite image of the scene—shown in the middle of Figure 1); and second, it computes the current position of the vehicle [see vector (a) in Figure 1] by correlating live video to a reference image. The mosaic is also the map used by the operator to specify the target location of the vehicle [see vector (b) in Figure 1]. It presents information in a familiar and intuitive fashion, which facilitates the specification of navigation commands. Finally, the combination of vectors (a) and (b) produces an error vector (c) used to control the vehicle.

Section B provides background on our previous research in vision-based station-keeping and mosaicking and explains the vision-processing used for navigation-from-video. Section C presents the user interface we have implemented for this task. Section D concludes with a discussion of our experimental results.

OTTER

The navigation-from-video capability has been demonstrated on OTTER (Oceanographic Technology Testbed for Engineering Research, see Figure 2 and [6]), a small underwater vehicle operated in a test tank at MBARI. OTTER is 2.1 meters long, 1 meter wide and 0.5 meters high, with a dry mass of 150 kg. Three pressure housings hold VME card cages, accelerometers, gyros, a gravity

---

[*] This paper will also be published in the proceedings of the IEEE Oceans '98 conference, Nice.

[†] Doctoral Candidate, Department of Electrical Engineering, huster@sun-valley.Stanford.EDU

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

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

[¶] Adjunct Research Engineer, MBARI

0-7803-5190-8/98/$10.00 ©1998 IEEE.

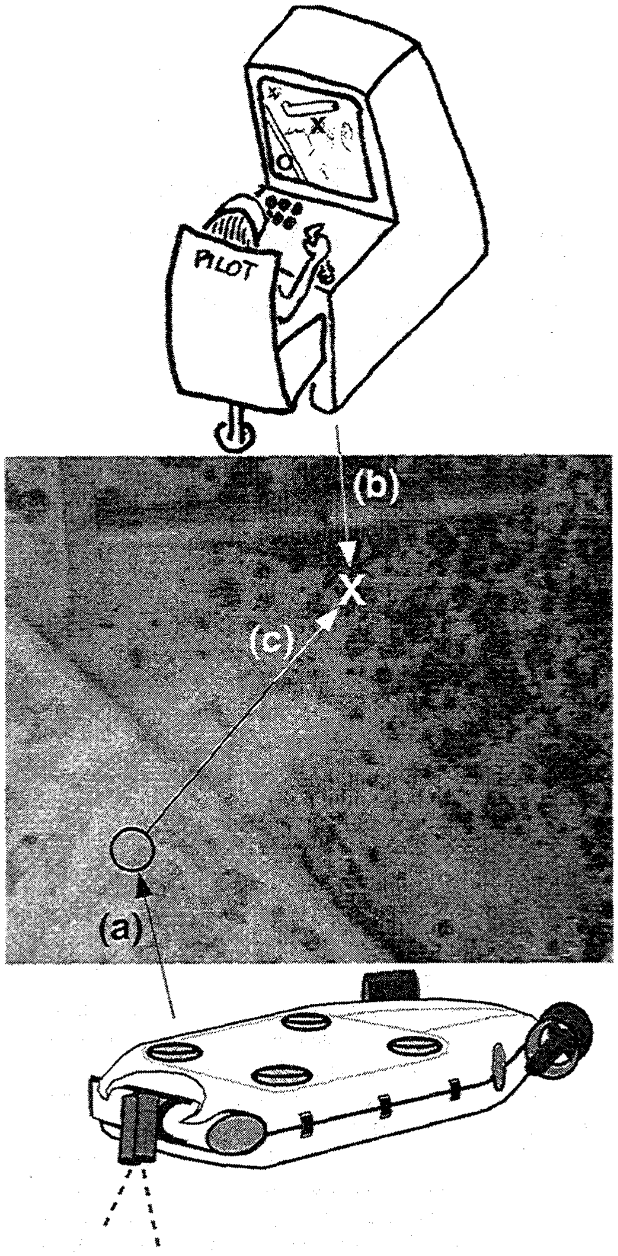

Figure 1: **Navigation Overview.** The video mosaic serves as a navigation map: (a) current location of the vehicle in mosaic; (b) target location specified on the mosaic by the pilot; (c) error vector used in vehicle control.

sensor, a fluxgate compass, a pressure sensor, all electronics, and NiCad batteries. It has 8 ducted thrusters to actuate all degrees of freedom, two CCD video cameras and a fiberglass fairing. A tether for Ethernet, power, and video, as well as remote vision processing, facilitates the development process.

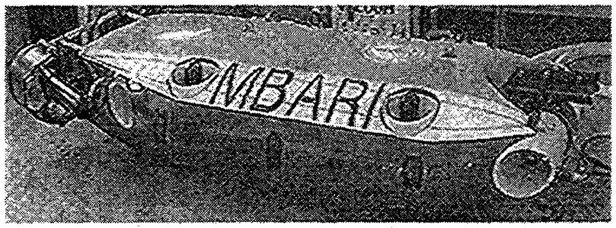

Figure 2: OTTER

B. Vision Processing

Navigation-from-video is an extension of vision-based station-keeping and mosaicking, which are described at the beginning of this section. The rest of this section describes the vision processing specific to navigation, including sources of error and related work.

Vision-Based Station-Keeping

Underwater vehicles are often required to hover, or hold station, relative to some feature in the environment. Automatic station-keeping combines both a local position sensor and control to regulate the vehicle position.

In previous work [4], vision has been presented as a drift-free local position sensor for robotic vehicles operating in unstructured, underwater environments. Our approach determines the motion parallel to the image plane of a standard, on-board video camera by correlating the current live image with a reference image and computing the offset between them. This vision-based sensor can be used together with depth and attitude sensors to obtain a drift-free measurement of position and orientation for an underwater vehicle. Because station-keeping implies a control system that keeps the vehicle close to its target location, every image is compared to the original reference image, and thus no bias accumulates over time.

We have demonstrated the effectiveness of this technology on both OTTER and *Ventana*, a remotely operated vehicle (ROV) used by MBARI for deep-ocean science missions [2].

Video Mosaicking

Video mosaics are collections of video images recorded at different locations and arranged such that corresponding edges overlap. In previous research [3], we have developed a method to build video mosaics in real time as an underwater vehicle moves over the ocean floor at constant altitude.

Video mosaics are particularly useful in underwater exploration because they enable images of larger regions. The strong attenuation and significant scattering of light in water limit the feasible distance between a camera and the ocean floor. Thus, it is not practical simply to snap an image from a higher altitude. However, by arranging many small video images, we are able to construct larger, more useful composite images of the ocean floor. This capability was also developed in the test tank using OTTER and demonstrated on *Ventana*, MBARI's ROV, in the ocean.

Vision Processing for Navigation

Navigation-from-vision is more general than station-keeping because the vehicle can be commanded to move to arbitrary positions. Thus, the vehicle is not confined to small regions and cannot use the same reference image throughout the navigation task. Furthermore, a mosaic needs to be generated and updated as the vehicle moves.

Before the vehicle moves away from its current reference image (i.e., while the current image and the reference still overlap and can be correlated), the vision processor must acquire a new reference. Each new reference image is stored and added to the mosaic that is displayed to the vehicle pilot. Figure 3 shows an example of this process.

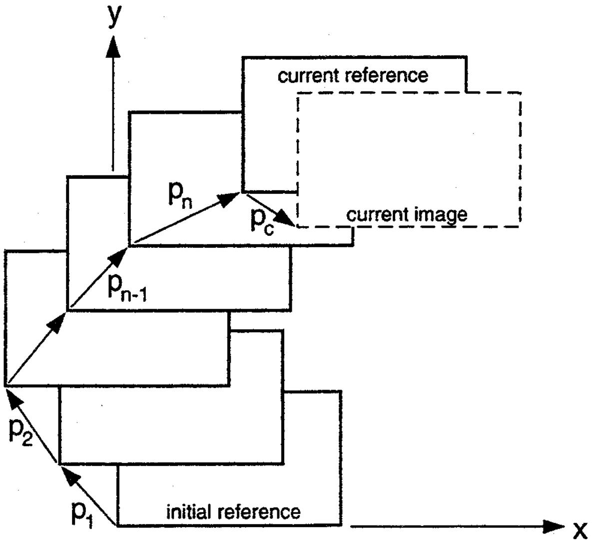

Figure 3: **Construction of Mosaic.** The rectangles indicate the reference images acquired as the vehicle moves. The overlap between references allows them to be correlated so that the offset between adjacent pairs (p_i) can be computed. The dashed rectangle shows the current image which is correlated with the current reference to produce the offset p_c.

The vision processor correlates each image to the current reference at the tail of the mosaic and determines the offset (p_c) between them. Then it adds this offset to the sum of offsets between the reference images p_i to obtain p, the position of the vehicle in a coordinate frame attached to the initial reference image.

$$p = \sum_{i=1}^{n} p_i + p_c$$

The vision processor produces position coordinates in two-dimensions—it cannot detect but will tolerate small changes in altitude and orientation—so the position is given as the location on a surface defined by the ocean floor.

Sources of Error

Every correlation between two video images is subject to two dominant sources of error: false matches and random noise. The image correlation step shifts the current image to all valid positions relative to the reference and evaluates the quality of the match at each of these offsets. Subsequently, it picks the offset that results in the best match. Occasionally, especially when the texture of the images is not rich enough, false matches are selected. This error occurs infrequently, but its magnitude can be large. Random noise, characterized by reasonably small magnitudes, is present in every correlation.

Systems that integrate optical flow—the offset between every subsequent pair of video images—to measure positions [5] are particularly susceptible to these types of errors. Our vision processor is much more robust to accumulating errors because it correlates current images to a reference that changes only when the vehicle moves out of the field of view of the current reference image. However, as the path length of the vehicle increases, the accumulated error in the position of the vehicle still grows without bounds. This weakness is inherent in any dead-reckoning sensor which integrates motion.

Related Work

In general, the drift inherent with dead-reckoning can only be eliminated by incorporating a measurement of position relative to a fixed point. However, in related research, we are developing algorithms to identify and remove the accumulation of error for mosaics generated when the vehicle path crosses over itself [1]. In this case, the current image can be correlated to a part of the mosaic that was generated at the beginning of a loop.

Correlating the beginning and the end of a loop has two consequences for the accuracy of the loop. First, any error that was accumulated while traveling along the loop no longer affects the accuracy of subsequent correlations. Second, the error that was accumulated along the loop can be identified and distributed back across the loop in order to increase the overall accuracy of the mosaic.

Because the ability to specify target positions and to compute the position of the vehicle during navigation are limited by the accuracy of the mosaic, the algorithms to increase accuracy in mosaics will also improve the quality of the navigation system.

C. Navigation Interface

We have developed a simple interface for navigating underwater vehicles: the pilot is presented with an evolving mosaic of the vehicle path and can specify new target locations simply by clicking the mouse on the mosaic (see Figure 4). An X marks the target position, and an O marks the current vehicle position. The pilot can specify new target positions at any time, regardless of whether previous moves have been completed.

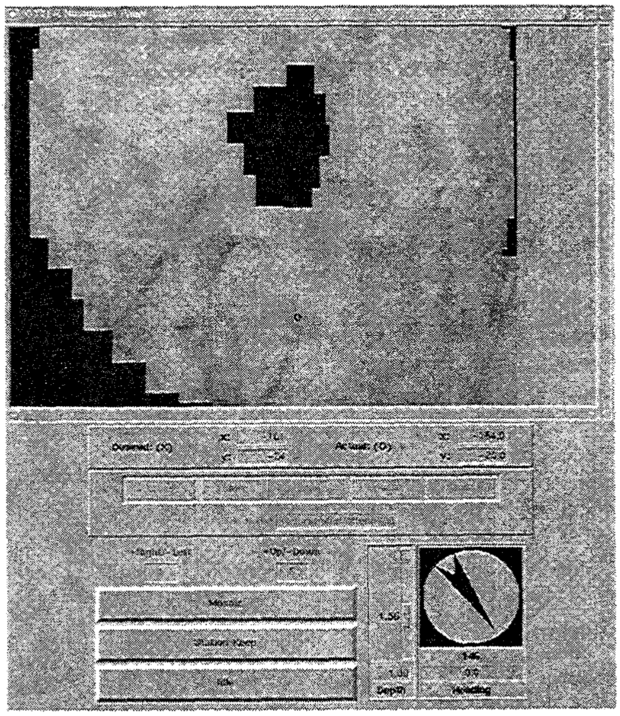

Figure 4: **Navigation Interface.** The top panel shows the evolving mosaic. Immediately below are the pixel coordinates of the desired and actual position. The compass on the bottom right shows the vehicle heading and the slider shows vehicle depth. The buttons on the lower left send additional commands to the vehicle.

In addition to the actual position of the vehicle, the interface shows the current heading and depth. Also available is a mosaicking command which tells the vehicle to scan a region systematically in order to build up a mosaic, a station-keeping command to stop the vehicle at its current position, and an idle command to stop controlling the vehicle.

Operational underwater vehicles tend to have many actuators and sensors which are all competing for attention on the user interface. Consequently, an experienced pilot familiar with the interface is required to monitor all of the sensors and alarms and to issue low-level commands. This navigation interface is designed to complement, and not replace, such a comprehensive interface. The intent is to provide a simple and intuitive high-level interface to direct the vehicle during vision-based navigation tasks.

Untethered vehicles are still very severely limited in capability because they lack the high-level perception and guidance of a human supervisor. A major challenge of autonomous vehicles is to use the limited bandwidth efficiently to maximize the high-level information presented to the user.

The interface requires only a low-bandwidth connection to the vehicle and is very tolerant of communication delays. For typical hovering vehicles, the mosaic update rate peaks at about one video image every 1–2 seconds and reduces to none while the vehicle is holding station. The vehicle also sends its current position within the mosaic at a low rate (e.g., 1 Hz). The interface sends only the new target positions to the vehicle at the rate at which they are generated by the pilot.

None of these data are time critical and any delay in the communication link merely reduces the speed at which the mission can be completed. For example, if the mosaic has not been updated for 5 seconds, the pilot may have to wait before issuing the next command, but in the meantime, the vehicle remains stable and continues to control to its current target position. Because of this tolerance to delay and bandwidth constraints, the interface can control an underwater vehicle over standard data networks, like the Internet.

D. Experimental Results

The navigation system was tested on OTTER in a test tank at MBARI. OTTER's camera is pointed downwards at a patterned shower curtain on the floor of the tank.

Figure 4 shows the mosaic that was generated from 36 video images as part of the navigation experiment. The path starts at the lower right corner and proceeds in a clockwise manner around a loop, finally crossing over its initial path. Figure 5 shows actual and desired positions for part of the loop.

Figure 6 shows the cross-over region in detail. Because of the accumulation of errors around the loop, a noticeable misalignment between the initial and final images has developed. Correlating between image 7 and image 35 shows the misalignment in X to be 26 pixels and in Y to be 50 pixels. This represents about 2.5% of

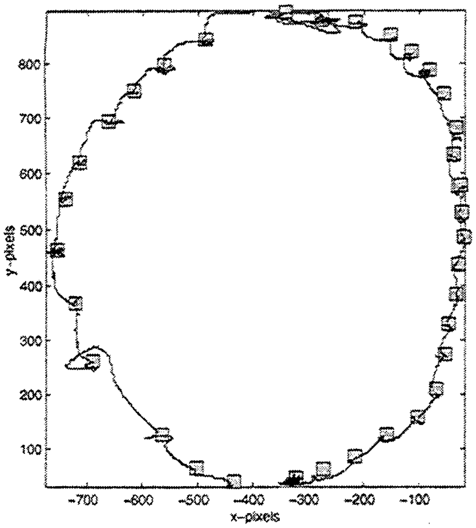

Figure 5: **Experimental Results.** The line indicates the actual position of OTTER in pixels; the squares are the desired position specified using the navigation interface.

the circumference of the loop, which is typical for this system.

We have also demonstrated the ability to pilot OTTER over the Internet from a site that is located more than 100 km from the test tank. Although the mission proceeds at a slower pace because of communication delays, the vehicle control remains stable. Even in the presence of delays, the navigation interface provides a useful and intuitive method of controlling an AUV or an ROV.

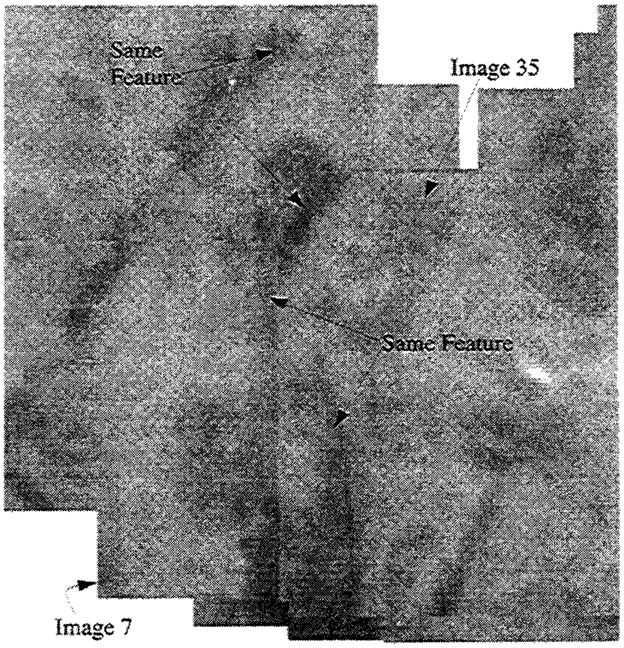

Figure 6: **Crossover Region of Mosaic.** Matching features shows the misalignment between the start (Image 7) and end (Image 35) of the loop.

E. Acknowledgments

This research is supported by the David and Lucile Packard Foundation under Grant #95-2777 with resources and support from MBARI.

References

[1] FLEISCHER, S. D., ROCK, S. M., AND BURTON, R. L. Development of a Vision-Based Global Position Sensor for Underwater Vehicles. In *Proceedings of the OCEANS 97 Conference* (Halifax, Nova Scotia, October 1997), MTS/IEEE.

[2] LEABOURNE, K. N., ROCK, S. M., FLEISCHER, S. D., AND BURTON, R. L. Station Keeping of an ROV Using Vision Technology. In *Proceedings of the OCEANS 97 Conference* (Halifax, Nova Scotia, October 1997), MTS/IEEE.

[3] 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.

[4] 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.

[5] NEGAHDARIPOUR, S., SHOKROLLAHI, A., FOX, J., AND ARORA, S. Improved methods for undersea optical stationkeeping. In *IEEE Int. Conf. on Robotics and Automation* (Sacramento, CA, April 1991), vol. 3, IEEE, pp. 2752–2758.

[6] WANG, H. H., ROCK, S. M., AND LEE, M. J. OTTER: The design and development of an intelligent underwater robot. *Autonomous Robots 3*, 2-3 (June-July 1996), 297–320.

More information, references, and papers are available at *http://arl.stanford.edu.*