REAL-TIME VISUAL MOSAICKING AND NAVIGATION OF THE USS Macon

Kristof Richmond¹, Stephen Rock¹,²
{kristof,rock}@stanford.edu

¹Aerospace Robotics Laboratory, Stanford University
Durand Bldg, Rm 028, 496 Lomita Mall, Stanford, CA 94035-4035

²Monterey Bay Aquarium Research Institute
7700 Sandholt Road, Moss Landing, CA 95039-9644

Abstract

The US Navy airship USS Macon was a 239 m dirigible aircraft carrier that went down off the Big Sur coast of California in 1935. Preliminary investigations of the wreck by the Monterey Bay Aquarium Research Institute (MBARI) and the US Navy in the early 1990s and in 2005 revealed two debris fields each about 50 m in diameter. In September 2006, a joint MBARI/NOAA expedition returned to the site for more extensive investigations with the ROV Tiburon aboard the R/V Western Flyer. The primary goal of the expedition was to document the wreck, and a primary output was to be a high-definition video survey of the two debris fields, with the ultimate goal of generating a high-resolution science-grade photomosaic.

In order to guarantee full visual coverage of the site with sufficient overlap of imagery for the output mosaic, traditional on-board positioning systems such as a Doppler velocity log (DVL) or a ship-to-vehicle ultra-short baseline (USBL) acoustic system were of insufficient accuracy for such a large area. Also, due to time and cost considerations, the deployment of precision positioning systems such as a long baseline (LBL) array at the site was considered infeasible. Instead, to provide precision positioning information and on-line feedback of the survey progress, a real-time visual mosaicking and navigation system, developed at the Stanford Aerospace Robotics Laboratory (ARL) in cooperation with MBARI, was employed.

The real-time visual mosaicking and navigation system provides high-precision, environment-relative vehicle positioning and control over the sea floor without the use of external positioning arrays. The primary components of the system are a DVL and a camera mounted on the vehicle. These two systems are complementary.

The DVL provides a continuous measurement of the vehicle’s position, even during periods of poor visibility. However, this position estimate alone is insufficient for large area surveys since it suffers a dead-reckoning drift. The vision system provides a second, independent means of estimating vehicle position by registering sequential images taken as the vehicle moves. Though such along-track registration is also subject to dead-reckoning drift, registering images at trajectory crossover points or where there is side-to-side overlap provides a means to eliminate that drift. Together, the information from these two components produce a map of the environment plus an unbiased estimate of the vehi-
cle’s position within that map. This in turn enables the vehicle to fly over a site guaranteeing complete coverage without using any external positioning system. Further, calculating the positions of images on the fly enables the real-time display of a navigation-grade mosaic of the visited area as an additional pilot aid. With some extensions, this system has the potential to be deployed on AUVs as well.

There are several aspects of this system which enable real-time capability. First, only the horizontal displacements between mosaic images (or tiles) are computed by the vision correlation algorithm. The remaining degrees of freedom are measured using available vehicle sensors (e.g. altimeter, compass and inclinometers). Further, the correlation is performed using a fast, texture-based algorithm. Second, the relative error between new tiles and their neighbors is kept low by demanding vehicle trajectories which maintain large amounts of overlap between successive passes. Third, 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, even with thousands of tiles. Finally, lighting variations and distortion of the images are not compensated for and no edge blending is performed. The result is a real-time mosaic that presents a rough, navigation-grade image of a site. At the Macon site, this system provided complete coverage of the two debris fields, with over 6000 images for each field.

1 Introduction

1.1 Expedition Background

The USS Macon was, along with its sister-ship the USS Akron, the largest US aircraft ever built. At 239 m, the lighter-than-air vessel was only 30 m shorter than the RMS Titanic. It was used in the early 1930s as a flying aircraft carrier by the US Navy, with a squadron of 5 Curtiss F9C-2 Sparrowhawk biplanes operating from a hanger in the belly of the airship. On 12 February 1935, while returning to its base at the Sunnyvale Naval Air Station (now Moffet Field) south of San Francisco, the Macon was struck by a sudden squall which tore off its top tail fin, damaging several of the helium bladders providing lift, and eventually leading to it hitting the ocean and sinking in around 400 m of water off the Big Sur coast of California. Eighty-one of the eighty-three crewmen aboard survived the accident [1]. After sinking, the lightweight aluminum-alloy frame composing the airship skeleton collapsed, leaving a relatively flat wreck site with only the more durable Sparrowhawk biplanes and the airship’s 12-cylinder Maybach engines sticking 1-2 m out of the sea floor.

The wreck site was first explored by joint US Navy and MBARI expeditions in the early 1990s and in 2005. Major artifacts such as the biplanes were identified using the ROV Ventana, and a sidescan sonar survey revealed two wreckage fields each approximately 50 m in diameter. In September 2006, an expedition headed by NOAA and MBARI returned to the wreck in order to document the site fully. As part of this documentation, a high-definition visual survey of the entirety of the two debris fields was to be performed, with the ultimate goal of creating high-resolution photomosaics from the captured video.

In order to guarantee full visual coverage of the site, precise vehicle positioning was required. To achieve the required level of precision while staying within the operational and costs constraints of the expedition required a novel underwater positioning technique. Traditional deployment of an LBL [2] would provided sufficient precision [12] but would have incurred costs in equipment, time and manpower to obtain, deploy, calibrate and retrieve which made it infeasible for this application. Lower-cost, self-contained positioning via DVL dead-reckoning would have provided an acceptable alternative from an operational point of view, but the ∼1% of distance traveled drift error in the position measurement [13, 14] would be too great.

To close the gap between precision and operational cost, the real-time visual mosaicking and navigation system, developed by the Stanford ARL in conjunction with MBARI, was employed. This system relies only on sensor systems already on board the vehicle, eliminating any additional operational expense. In addition, vision provides a direct measure of vehicle position relative to the seafloor, eliminating drift relative to the objects of interest. Finally, as the vehicle maneuvers, the system is able to provide a real-time overview of the surroundings, which can be used for planning and observation purposes. It should be noted that the goal of this system is not to produce scientific-grade high-resolution photomosaics, but rather to provide sufficiently precise environment-relative positioning information to guarantee complete coverage for a visual survey.

1.2 Underwater visual mosaicking

Due to the limited transmission of electromagnetic radiation—and in particular light—under water, obtaining large-scale visual overviews of the sea floor is not possible with a single image. Rather, many images—taken at a sufficiently low altitude that reflected light is not significantly attenuated—must be combined together to get a view of large-scale structure. This process is known as visual mosaicking, and has long been recognized as a significant tool for underwater explo-
The attenuation of electromagnetic radiation by water also prevents precision radio-frequency positioning systems, such as differential GPS, from being employed as an underwater sensor. While acoustic methods (LBL, SBL, USBL) have been developed for navigation, several authors have noted that the rich, environment-relative information coming from vision can be used to provide vehicle positioning, using techniques related to the post-processed underwater mosaicking systems. Marks and Fleischer [6, 7] performed early investigations of fast image registration and small-scale real-time mosaicking techniques dealing with some of the limitations of the underwater environment such as moving lighting and marine snow and incorporating information from complementary standard underwater sensors such as inclinometers, compass and altimeter. The high-speed image acquisition and registration framework they developed forms the core of the system described here. Their work has also been extended by Garcia, et al. [8] into a Kalman filter framework with detailed vehicle models and demonstrated in a test tank.

Gracias, et al. [9] achieved success in shallow waters with ambient lighting using a previously acquired and processed mosaic by which to navigate. Eustice, et al. [10] performed a post-processed demonstration of a navigation technique fusing DVL and vision measurements to determine vehicle position and produce a 3-D map of the imaged surface using data collected at the RMS Titanic. Madjidi and Negahdaripour [11] investigated a purely vision-based method for determining full vehicle pose and a globally aligned mosaic, with potential near-real-time performance, which they demonstrated on a small (∼3 m × 3 m) area mosaic collected in the field.

2 System description

The visual mosaicking and navigation system used at the Macon site has been described previously [15, 16]. To enable the robust, real-time performance which allowed this system to be successfully deployed in the field, several design decisions were of critical importance.

To improve overall robustness, the system consists of two primary, complementary components: a DVL and a downward-looking video camera mounted next to each other on the ROV (see Figure 2). The DVL provides a continuous measurement of the ROV position, even during periods of poor visibility. Real-time measurements from the vision subsystem are used to overcome the DVL drift relative to the local environment by registering images at trajectory cross-over points or where there is side-to-side overlap.

![Figure 2: ROV Tiburon set up to perform mosaicking for the Macon expedition. The down-looking HDTV camera is indicated by the arrow. The DVL was mounted on the frame next to the camera, and outriggers provided an even light field with reduced backscatter.](image-url)
held constant to eliminate the need to rotate images (reducing the visual processing load) and high-aspect-ratio (16:9) video imagery was used. To take advantage of the greater image width and thus amount of overlap available side-to-side, while maintaining the constant heading, it was decided to use a lawnmower pattern with the horizontal image dimension oriented across the direction of travel.

Third, to compute the positions of tiles in the mosaic, a fast, flexible information filter framework is employed. This filter, based on the formulation of Lu and Milios [19], provides the real-time global alignment— or optimal estimates of tile locations—for mosaics containing thousands of tiles. The position $p_i$ of each tile is taken as a variable to be estimated. The concatenation of these positions forms the state vector of the mosaic:

$$x = \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_N \end{bmatrix}, \quad x \sim \mathcal{N} (\mu_x, \Sigma_x).$$

A measurement $z_{ij}$ of displacement between tiles from the vision registration or from the DVL forms a constraint between the individual tile positions referred to as a link (see e.g. Figure 3). The strength of each link is given by the variance of the measurement:

$$z_{ij} = p_j - p_i + \nu = C x + \nu, \quad \nu \sim \mathcal{N} (0, \Sigma_{z_{ij}}).$$

The information filter [20] tracks the information state $\xi$ and the information matrix $\Omega$, where

$$\Omega = \Sigma_x^{-1}, \quad \xi = \Omega x.$$

The global alignment system continues the assumptions made by the vision registration that only the two translational components of the vehicle motion parallel to the bottom plane need to be computed, so that

$$p_i = \begin{bmatrix} x_i \\ y_i \end{bmatrix}.$$

The remaining components of the 6-DOF vehicle state take their 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. Thus, the filter measurement update is a simple incrementing of the appropriate elements of $\xi$ and $\Omega$:

$$\xi' = \xi + C^T \Sigma_{z_{ij}}^{-1} z_{ij},$$
\[ \mathbf{Q} = \mathbf{Q} + C^T \mathbf{S}^{-1} C. \]

To recover the state estimate of all mosaic tile positions \( \hat{x} \) requires solving the sparse symmetric positive definite system

\[ \mathbf{Q} \hat{x} = \xi \]

which can be solved for state vectors tracking thousands of tiles on standard PC hardware in less than a second with standard sparse solvers. The recovery of \( \hat{x} \) is not necessary for the filter propagation, and can thus be performed on an as-needed basis given the available processing power.

Finally, when presenting the resulting mosaic, lighting variations and distortion of the images are not compensated for and no edge blending is performed. The output of the visual mosaicking and navigation system is a real-time mosaic that presents a navigation-grade overview of a site and a real-time estimate of the position of vehicle, which in turn enables the vehicle to fly over a site with enough positioning precision to guarantee complete coverage. In addition, the real-time mosaic provides rich, intuitive feedback and environmental awareness to human users and serves as an additional pilot aid.

### 3 USS Macon expedition

The NOAA/MBARI expedition to the USS Macon took place from 18-22 September 2006. Three of the five days of the expedition were devoted to gathering the high-definition video for the photomosaics. The real-time visual mosaicking and navigation system was successfully able to completely image both of the known debris fields. Figure 6 shows one of the navigation-grade real-time mosaics gathered by the system, covering about one quarter of the total area of debris.

#### 3.1 Challenges

In order to ensure that a potential failure of the automatic search for side-to-side registration would not endanger the success of the mosaicking task, a manual backup system was implemented. This consisted of a user interactively marking a pair of images toward the ends of adjacent passes which were observed to mutually contain an object or feature. The side-to-side registration process could then be forced to attempt registration of these pairs with a very high likelihood of success, ensuring that DVL drift accumulated during a pass would be eliminated. This procedure was implemented for most runs at the Macon site. When the collected data were re-run while disregarding the manually matched pairs, the automatic search was found to succeed on its own for 98% of passes.

While the DVL and vision provide a complimentary set of measurements which significantly improved the robustness of the positioning system to failures generally found in the field, the Macon site provided one impediment which the combined sensors were not able to overcome. It was found that at slack tide, large schools of fish were attracted to the vehicle lights and obscured the seafloor with their bodies and the dust they stirred up (see Figure 4). The concentrations were dense enough that DVL bottom-track was often corrupted by returns from the moving bodies, so that the drift error went up significantly and vehicle control was affected.
Figure 6: Navigation-grade real-time mosaic A2, covering slightly less than half of Field A of the Macon wreck containing the main hangar area with remains of four Sparrowhawk biplanes and various other debris. The area where a gap in overlap appears at the lower right was not the result of navigation error, but rather a deliberate attempt to test the ability of the system to close large loops. The area was imaged in a subsequent run.
Figure 7: Real-time mosaic superimposed on high-resolution bathymetry obtained from mosaicking system. The Sparrowhawk biplanes can be seen in the slightly raised areas toward the center of the image. Vertical scale is exaggerated by a factor of two in this view.

Table 1: Statistics for the real-time mosaics gathered at the Macon wreck site.

<table>
<thead>
<tr>
<th>Mosaic</th>
<th>No. of Images</th>
<th>Approx. Area (m²)</th>
<th>Path Length (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>1120</td>
<td>450</td>
<td>1120</td>
</tr>
<tr>
<td>A2</td>
<td>2966</td>
<td>1370</td>
<td>1800</td>
</tr>
<tr>
<td>A3</td>
<td>3361</td>
<td>1380</td>
<td>2320</td>
</tr>
<tr>
<td>B1</td>
<td>1451</td>
<td>650</td>
<td>1310</td>
</tr>
<tr>
<td>B2</td>
<td>4435</td>
<td>2100</td>
<td>3210</td>
</tr>
</tbody>
</table>

Due to these tidally-synchronized fish events, mosaicking operations had to be divided into ~4 hr mosaicking periods interspersed with ~2 hr periods of rest or alternate operations at slack tide. This also limited the area which could be mosaicked at one time, so that the final mosaics of the two fields are composed of sub-images which have been manually pieced together. Table 1 shows the sizes of the real-time mosaics gathered at the Macon site.

While the terrain on which the Macon wreck came to rest is relatively flat, the site did present some challenges in relief. In particular, the Sparrowhawk biplanes, with their “sky hooks” attached to the top of the aircraft and used to retrieve them into the airship in mid-flight, presented a collision danger to the vehicle. This required that in the vicinity of the biplane remains, the vehicle fly at a much higher altitude from the bottom than was optimal elsewhere in the debris field, about 4.5 m vs. 1.5 m. The vision system was able to successfully handle the threefold change in altitude and the resulting changes in scene lighting and scale. In addition, the real-time mosaic proved useful in providing situational awareness of the planes’ locations, as obstacle-avoidance cameras mounted on the vehicle proved ineffectual due to poor lighting and restricted field of view.

### 3.2 Extensions and derivative results

While this work has thus far been deployed on tethered ROVs, visual navigation holds great promise for expanding the capabilities of AUVs as well. Exploration and eventually intervention capabilities will benefit from the rich, high-precision, environment-relative measurements provided by vision. A detailed visual map, such as that produced by this system, provides not only an operationally and scientifically interesting data set, but can also serve as a basis for automated real-time path and task planning.

While the system presented here does assume that it is working over a planar surface, fairly large changes in relief can be accommodated as long as they are sufficiently gradual. In fact, the precise translational positions returned by the system can be combined with the vehicle depth and altitude measurements to obtain high-resolution bathymetry over which the visual mosaic can be laid to form a rough three-dimensional view of the seafloor. Since this depth and altitude information is already available in real time, online implementation of such a view presents itself as a possibility, given sufficiently capable display facilities. Figure 7 shows an example of this type of 3D visualization applied to data.
Figure 8: High-resolution post-processed photomosaic of the hangar portion of the Macon wreck showing the remains of two of the Sparrowhawk airplanes on board the airship when it went down. The high-resolution mosaic is based off a $\sim 8 \text{ m} \times 12 \text{ m}$ portion of the real-time mosaic, containing $\sim 100$ images.

from the Macon.

The large amount of high-resolution imagery gathered is currently being processed into the final mosaic product. The post-processing is being performed using the system developed by Singh, et al. at WHOI [3], with slight modifications to incorporate the navigation solution from the real-time mosaicking system. Figure 8 shows some preliminary results of the high-resolution mosaic. Since the post-processed mosaic optimizes over the full set of possible camera motions, it can be used as a sort of truth measurement for the real-time navigation solution. For the areas processed thus far, the positions computed by the real-time navigation system agrees to within $\sim 10 \text{ cm}$ of the post-processed solution, confirming that the desired degree of precision has been achieved.

4 Conclusion

This paper has presented a large-scale operational demonstration of a vision-based underwater navigation system. Several aspects of the system make it robust and fast enough for implementation in the field. The use of a DVL and a video camera allows the system to take advantage of the strengths of both sensors to increase robustness. Introducing vision as a sensor also introduces a large amount of information to be processed, so that several aspects of the system had to be tuned to achieve real-time performance. The system was successfully deployed on an archaeological expedition to image completely two large debris fields of the wreck of the USS Macon.

Acknowledgments

The authors would like to thank the Monterey Bay Aquarium Research Institute for its support in performing this research. Special thanks go to the pilots of both the Ventana and Tiburon ROVs for their help in developing and testing the system. Thanks also to NOAA for sponsoring the expedition to explore the Macon crash site, and to Hanumant Singh and the Woods Hole Oceanographic Institute for use of their high-resolution post-processed mosaicking code.

References


