Stereo Vision–Based Navigation for Autonomous Surface Vessels

Article *in* Journal of Field Robotics · January 2011

DOI: 10.1002/rob.20380 · Source: DBLP

4 authors, including:

Terry Huntsberger
Huntington Ingalls Industries, Inc.
183 PUBLICATIONS **3,646** CITATIONS

SEE PROFILE

Hrand Aghazarian
NASA
57 PUBLICATIONS **1,734** CITATIONS

SEE PROFILE

David C. Trotz
California Institute of Technology
1 PUBLICATION **118** CITATIONS

SEE PROFILE

Stereo Vision–Based Navigation for Autonomous Surface Vessels

Terry Huntsberger and Hrand Aghazarian
Mobility and Robotic Systems Section, Jet Propulsion Laboratory, Pasadena, California 91109
e-mail: Terry.Huntsberger@jpl.nasa.gov, haghazar@jpl.nasa.gov
Andrew Howard
Space Exploration Technologies, 1 Rocket Road, Hawthorne, California 90250
e-mail: Andrew.B.Howard@gmail.com
David C. Trotz
11611 N. Howard Court, Spokane, Washington 99218
e-mail: David.Trotz@gmail.com

Received 27 January 2010; accepted 31 October 2010

This paper describes a stereo vision–based system for autonomous navigation in maritime environments. The system consists of two key components. The Hammerhead vision system detects geometric hazards (i.e., objects above the waterline) and generates both grid-based hazard maps and discrete contact lists (objects with position and velocity). The R4SA (robust, real-time, reconfigurable, robotic system architecture) control system uses these inputs to implement sensor-based navigation behaviors, including static obstacle avoidance and dynamic target following. As far as the published literature is concerned, this stereo vision–based system is the first fielded system that is tailored for high-speed, autonomous maritime operation on smaller boats. In this paper, we present a description and experimental analysis of the Hammerhead vision system, along with key elements of the R4SA control system. We describe the integration of these systems onto a number of high-speed unmanned surface vessels and present experimental results for the combined vision-based navigation system.

1. INTRODUCTION

The United States Navy is currently developing uncrewed surface vessels (two examples are shown in Figure 1) for a variety of missions, including patrol, reconnaissance, and force protection. These vessels must have a very high degree of mission-level autonomy, with capabilities that include waypoint navigation, hazard avoidance, and adherence to maritime "rules of the road." In support of this effort, the Jet Propulsion Laboratory (JPL) has developed a complete autonomy suite for unmanned surface vessels (USVs). The CARACaS (Control Architecture for Robotic Agent Command and Sensing) suite (Huntsberger, Aghazarian, Castano, Woodward, Padgett, et al., 2008) includes support for high-level mission planning and resource management, midlevel navigation behaviors, and low-level sensing and control. A key aspect of safe autonomy on the water is the hazard avoidance sensing and algorithms that must be capable of robust performance in a wide variety of water surface and environmental conditions and at speeds in the range of up to 35 knots. In this paper, we describe two components of this suite: Hammerhead, a vision-based system for detecting and tracking both static and dynamic objects from a moving vessel; and the R4SA (robust, real-time, reconfigurable, robotic system architecture) framework, which implements fundamental navigation behaviors such as obstacle avoidance and target following. As far as we know, the Hammerhead/R4SA system is the first fielded system that has been designed and tested on the water for high-speed, autonomous small boats (13–14 m and smaller).

The Hammerhead system is designed to detect geometric hazards on water; i.e., objects that protrude above the waterline. It consists of both hardware and software components. The hardware component is a sensor head containing four cameras arranged in left-facing and right-facing stereo pairs. All four cameras lie on the same baseline, which allows Hammerhead to achieve a wide field of view without increasing the footprint of the sensor head. The software component is a multistage processing algorithm that takes as input images from the cameras and produces two outputs: a probabilistic hazard map and a list of contacts (i.e., discrete objects) with estimated speed and heading. The hazard map is a discrete grid of cells that are labeled as hazard/contact and nonhazard based on models derived from the stereo processing. Labels also include identification of the hazards/contacts derived from ENC (electronic nautical charts) of the area (e.g., buoys, channel markers) or from a model-based analysis of the objects. As currently configured, Hammerhead runs at 4 Hz on a Core2

Journal of Field Robotics 28(1), 3–18 (2011) © 2010 Wiley Periodicals, Inc.
View this article online at wileyonlinelibrary.com • DOI: 10.1002/rob.20380

Figure 1. Two members of the autonomous boat fleet: HTF (left) and CMV (right). The Hammerhead stereo bar is mounted on the lower portion of the arch, with the SAVAnT 360-deg camera array directly above. The CMV also has radar and LIDAR sensors.

Quad processor and has an effective range of about 200 m over a 90-deg field of view. The longer look-ahead range, due to the unique four-camera design of the system, helps to mitigate the 4-Hz update rate (corresponds to about a $^1/_4$ boat length at top speed). This paper describes the Hammerhead stereo vision system and presents both quantitative and qualitative assessments of system performance.

R4SA is an adaptable, real-time, embedded software architecture that monitors and controls the boat actuators (including engine, throttle, and rudder) while also implementing core behaviors such as station keeping, waypoint navigation, obstacle avoidance, right of way, target intercept, and target following. The R4SA system has been integrated into a small fleet of high-speed autonomous boats (six at latest count) and has seen testing under a wide variety of conditions and scenarios. The probabilistic gridded hazard map produced by the Hammerhead system is used as input to an onboard path planner for safe navigation. This paper describes two R4SA navigation behaviors that make use of visual data: static obstacle avoidance and dynamic target following. We present results for these behaviors from integrated field tests.

The paper is structured as follows. In Section 2, we briefly review the related literature on sensor-based maritime autonomy. In Section 3, we describe the system hardware and software architecture and discuss some of the practical difficulties of working with small boats in maritime environments. In Sections 4 and 5, we describe the software components of the Hammerhead stereo vision system and present experimental results from on-water tests. In Section 6, we describe the specific algorithms for obstacle avoidance and following that are implemented in R4SA and show how they interact with the Hammerhead system. Finally, we present results from integrated on-water field tests and discuss some of the challenges of robust navigation in maritime environments.

2. RELATED WORK

Unlike terrain navigation by unmanned ground vehicles, autonomous maritime navigation is a relatively new field of investigation, and as such, there are correspondingly few fielded systems capable of sensor-based hazard detection and avoidance. Stereo sensing techniques for ground vehicles (Goldberg, Maimone, & Matthies, 2002; Matthies, 1989) are very mature and have been used on various Department of Defense and DARPA–sponsored projects and the Mars Exploration Rovers (MER) for autonomous navigation. In some sense, navigation over benign terrain is simpler than that over a dynamic water surface. On the other hand, the wheel/terrain interaction dynamics for terrestrial navigation have not been totally characterized because sensing of soil characteristics ahead of the vehicle is just now being investigated (Angelova, Matthies, Helmick, & Perona, 2007; Helmick, Angelova, & Matthies, 2009; Iagnemma, Shibly, & Dubowsky, 2002; Talukder, Manduchi, Rankin, Owens, Matthies, et al., 2002).

In the maritime navigation domain, Larson, Ebken, and Bruch (2006) have reported a behavior-based hazard avoidance system for USVs that combines deliberative path planning with reactive response to close-in dynamic obstacles. Their system uses digital nautical charts (DNC) for the initial long-range path planning coupled with a passive stereo system developed at JPL for the reactive control. The Hammerhead system is a second-generation version of the system used by Larson et al. and has over twice the spatial resolution. In addition, there are algorithms in the new system for glare reduction, handling of reflections on the water, and noise mitigation in low-light conditions. In more recent work (Larson, Bruch, Halterman, Rogers, & Webster, 2007), these authors also describe monocular and radar-based systems for obstacle detection. The monocular system uses the object distance from the horizon as a range cue; the authors note the difficulty in reliably finding both the horizon and the obstacles under a wide range of illumination conditions. The radar system, in contrast, generates reliable detections but has a relatively slow scanning rate that makes accurate data registration difficult on high-speed vessels. Another recently announced commercial system, the Interceptor USV, a joint effort by AAI, Marine Robotic Vessels International, and Sea Robotics Company, is a 6.5-m craft that has optional hazard avoidance

capabilities, but no details about method or fielded trials are openly available.

Laser range finders have also been employed for maritime hazard detection by a number of teams. Ruiz and Granja (2009) describe a system designed to assist human pilots in inland waterways. It adapts a commercial Riegl LIDAR, converting it to a two-axis scanner, and processes the resultant two-dimensional (2D) range image to detect discrete contacts. The maximum effective range is between 300 and 500 m, but the authors note some difficulties using the system on moving contacts due to the long scanning times (on the order of several seconds). Bandyopadhyay, Sarcione, and Hover (2009) describe a reactive obstacle avoidance approach for crowded harbor environments, using a SICK LIDAR as the primary sensor. The main focus of this work is the navigation controller, but the authors report that the SICK is relatively ineffective at ranges greater than 20 m due to the effects of wave motion.

In the monocular camera domain, Snyder, Morris, Haley, Collins, and Okerholm (2004) successfully demonstrated the components needed for autonomous navigation in harbor and riverine environments. Their system uses six cameras arranged as a 360-deg color sensor coupled with sky/sea/shoreline segmentation, optic flow, and structural model techniques to determine the relative position of obstacles and safe paths. More recently, JPL has demonstrated the 360-deg SAVAnT camera system (Wolf, Assad, Kuwata, Howard, Aghazarian, et al., 2010), which is capable of detecting specific classes of objects at ranges greater than 1,000 m. Although it is not primarily designed as an obstacle detection system, SAVAnT shows the promise of high-resolution, 360-deg camera arrays for full situational awareness.

Most recently, Spatial Integrated Systems, Inc. (Elkins, Sellers, & Monarch, 2010; Zhang, Zhuang, Elkins, Simon, Gore, et al., 2009), in a collaboration with JPL, D.H. Wagner Associates, and the Advanced Research Laboratory at Penn State, has demonstrated fused map generation for hazard avoidance and situational awareness using the JPL Hammerhead stereo, the JPL 360-deg SAVAnT camera, Ka-band radar, and a Velodyne LIDAR system (integrated sensor package seen on arch of the CMV in the right-hand image of Figure 1) during the Trident Warrior 2009 exercises in Norfolk, Virginia. The group demonstrated cross cueing of the Ka-band radar and the JPL 360-deg SAVAnT systems for contact identification and tracking (Wolf et al., 2010).

Comparing the dominant sensor modalities used for autonomous surface navigation (i.e., radar, LIDAR, monocular vision, and stereo vision) we make the following observations:

- Radar has by far the longest range but is limited by the mechanical scanning rate. Off-the-shelf maritime X-band radars also have difficulty processing short-range returns (i.e., less than 100 m), but this limitation could be overcome with better signal processing and/or custom solutions. Ka-band radars offer the best compromise with relative ranges from 30 m out to ∼3 km, but identification of vessel type is difficult using purely radar.
- LIDAR has reasonable range and good range resolution but like radar is limited by the mechanical scanning rate. In practice, the effective range is also limited by the relatively low angular resolution, which results in very few returns from small targets (such as boats) at longer ranges.
- Monocular vision has excellent angular resolution and has no need for mechanical scanning but is limited by the ability of current vision algorithms to reliably detect all types of obstacles in all lighting conditions. Range resolution is also relatively poor and relies on finding the horizon line in the image (either visually or through inertial sensing).
- Stereo vision also has excellent angular resolution and no need for mechanical stabilization. As we will show in this paper, it can reliably detect objects that protrude above the waterline. The primary limitation is range resolution, which is constrained by the camera baseline and image resolution.

Autonomous USVs will almost certainly require a combination of these sensor types to achieve reliable performance in all conditions. The next section provides a system overview of the boats, autonomy system, stereo system, and control architecture of the fielded components.

3. SYSTEM OVERVIEW

The U.S. Navy maintains a collection of autonomous boats at Ft. Monroe, Virginia. These boats are used for the development and testing of autonomy algorithms in the adjacent harbor, river, and offshore waters. Safety officers are onboard during the testing of the algorithms with full override capabilities because the test zones are not in restricted waters. A high-level diagram of the software/hardware system onboard the boats is shown in Figure 2. Because there is a wide variety of boats maintained at Ft. Monroe with different interfaces (including legacy systems), an autonomy kernel must be able to interface at any level from the throttle/rudder actuators all the way up to running on top of an autohelm system.

The CARACaS autonomy kernel runs under the embedded real-time R4SA that is derived from the flight technology used for the MER ATLO (assembly, test, launch, and operations) process prior to launch in 2003. R4SA is currently configured to run under the real-time operating systems VxWorks or QNX. Mission-level goals are usually fed into the autonomy kernel from a command and control (C2) station on shore or onboard another boat. Ideally, only minimal interaction with the autonomy kernel onboard the USV would be needed, other than periodic updates of situational awareness information, after it accepts the mission.

Figure 2. High-level system overview of a general maritime system. The autonomy kernel is responsible for keeping track of the world state, including the progress of the mission. External situational awareness information is fused with the onboard sensing (diagram courtesy of Eric Hansen, NAVSEA Carderock Division).

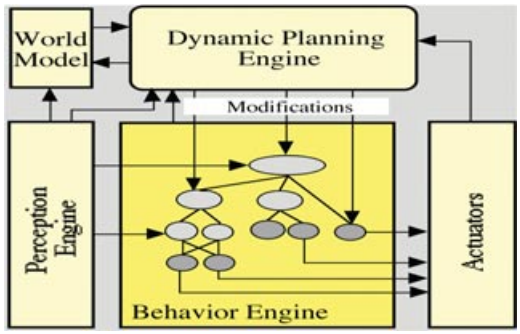

Figure 3. Block diagram of CARACaS. The network in the behavior engine is built from primitive (dark gray) and composite (light gray) behaviors. The dynamic planning engine interacts with the network at both the primitive and composite behavior levels.

The following subsections give brief descriptions of the autonomy and sensing system components.

3.1. Platforms

The Hammerhead/R4SA system has been fielded on six high-speed patrol craft, two of which are shown in Figure 1. The boats are fully drive-by-wire, using a standard maritime controller-area network (CAN) bus to interface with engines, throttle, and rudder. The commands to the throttle usually consist of percent of total range and to the rudder a desired angle. Some of the onboard autohelm systems have embedded closed-loop control to maintain stability around the desired setting, but the majority run in "override" mode when under autonomous control (rely on the autonomy kernel to maintain a steady state). An auxiliary generator powers a climate-controlled electronics box, which contains the servers running Hammerhead, R4SA, and other CARACaS autonomy components. A sensor arch (see Figure 1) supports a variety of sensing components, including the JPL Hammerhead camera bar and the JPL SAVAnT 360-deg camera array. A Crossbow 440 inertial navigation system (INS) unit provides position and heading information, with a typical accuracy of 3–5 m in position [using wide area augmentation system (WAAS)-enabled global positioning system (GPS)] and 0.2 deg in heading.

3.2. CARACaS Autonomy Engine

Overall vessel supervision is performed by CARACaS, which includes reasoning, planning, and perceptual and behavioral components, all tightly coupled with a world model (Hansen, Huntsberger, & Elkins, 2006). CARACaS (see Figure 3) explicitly addresses two key components of autonomy control systems: deterministic reaction to unanticipated occurrences and on-the-fly mission subtask replanning. All of the underlying behaviors in the behavior engine portion of CARACaS are built with finite state machines to guarantee appropriate responses within a predictable time frame through the JPL-developed R4SA framework. The world model maintains all vehicle state information, as well as the mission-level goals and current interactions with other vehicles.

3.3. Hammerhead Stereo System

Figure 4 shows a block diagram of the Hammerhead system. At the coarsest level, images from the stereo cameras are captured, logged, and processed by the stereo server application, which outputs a local hazard map and a list of contacts. These are transmitted over a wired network to the R4SA system, where they support hazard avoidance and tracking/following behaviors. A secondary, low-bandwidth data channel transmits telemetry data, images, and maps to a remote console. This channel uses a lossy user datagram protocol (UDP)-based protocol that tolerates very poor wireless network connections and allows remote monitoring and control of the Hammerhead system from a chase boat or shore installation.

Image data can be logged to a removable disk drive for later playback and analysis. In "replay" mode, the server operates exactly as it does in "live" mode, with the exception that images and INS data are read from the log rather than the physical devices. This mode facilitates off-board analysis of the Hammerhead system, along with integrated testing with other components. Note that achieving reliable data logging in this environment is a nontrivial problem; the boats experience very high accelerations, such that spinning disk drives can become unresponsive for many seconds at a time. The Hammerhead system buffers several seconds worth of image data in system memory before logging them to disk.

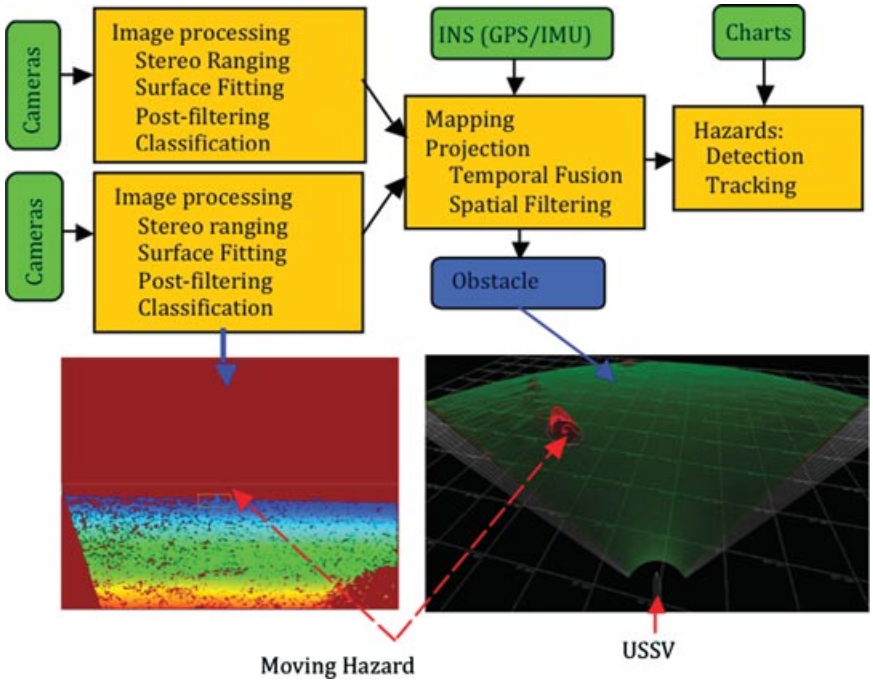

Figure 4. Block diagram of the Hammerhead system. Images are captured from two stereo camera pairs and processed through multiple stages to produce a hazard map and discrete contact list.

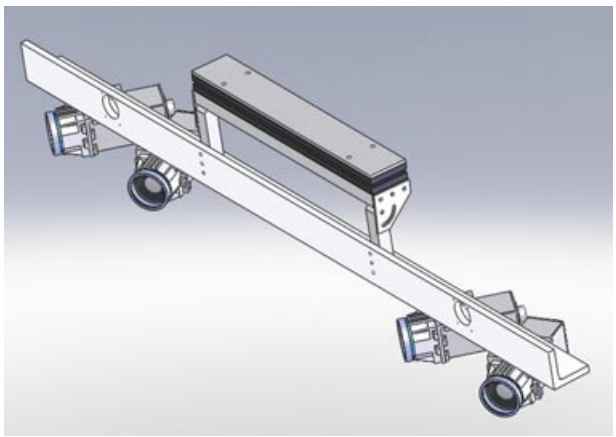

Figure 5. Hammerhead sensor head. The cameras are arranged in two stereo pairs (left facing and right facing) with a 1-m baseline and a combined field of view of approximately 100 deg. The cameras have monochrome CCDs with a resolution of 1,280 × 960 pixels and 60-deg horizontal field of view.

The stereo camera arrangement (see Figure 5) is somewhat unusual and reflects an engineering compromise between resolution, field of view, and form factor. With a single stereo pair, it is difficult to achieve both a large field of view (required for effective navigation) and a small angular resolution (required for long-range detection). In contrast, multiple stereo pairs can provide the required field of view and angular resolution but are unwieldy and difficult to integrate when mounted on separate fixtures in the traditional manner. The compromise, then, is to use two stereo pairs mounted to the same fixture. The cameras share a common baseline, but one pair faces left and the other faces right. This arrangement provides good resolution and field of view in a compact form factor, at the price of having has less-than-ideal geometry for stereo ranging. The final sensor configuration has two sets of cameras mounted on a common fixture with a 1-m baseline and 100-deg combined field of view. The cameras have monochrome charge-coupled devices (CCDs) with a resolution of 1,280 × 960 pixels and individual fields of view of approximately 60 deg.

Note that obtaining good quality images of both the water surface and the objects thereon can be challenging. Environmental factors such as low sun angles and specular reflection from the water surface can conspire to overwhelm the cameras' built-in autoexposure control, and high boat accelerations can lead to motion blur in the images. Consequently, Hammerhead includes an exposure control algorithm that compensates for these factors through adaptive gain control and shutter speed.

3.4. R4SA Control System

R4SA is an adaptable, real-time, embedded software architecture that has been in development at JPL for 14 years. It

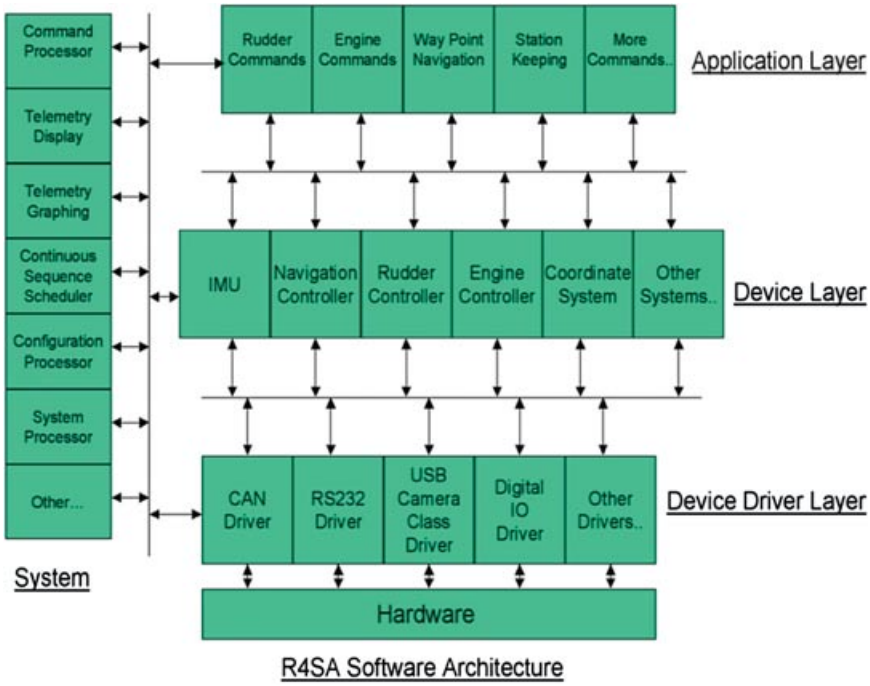

Figure 6. Block diagram of the R4SA control system for autonomous surface and underwater vehicle.

adheres to the design principles of low coupling, high cohesion, and testability, while supporting basic properties such as responsiveness, timeliness, schedulability, predictability, and determinism. Figure 6 shows a block diagram of the R4SA architecture as it is applied to USVs. The four layers of control are as follows:

- The device driver layer (DDL), which controls hardware buses (serial, CAN, Ethernet, etc.)
- The device layer (DL), which implements controllers and health monitoring for individual devices such as engines, throttle, and rudder
- The application layer (APL), which implements behaviors such as waypoint navigation, station keeping, hazard avoidance, and target following
- The system layer, which provides a scheduler for synchronization of tasks and a sequencer that executes commands received from external or internal sources

R4SA also provides interfaces for remote monitoring, control, and configuration and runs on top of a standard real-time operating system (QNX). All of the code is written in ASCII C. The majority of development for maritime autonomy is done within the APL, and the developer is sheltered from the underlying details of the actuators, etc. There is an onboard command dictionary that is the baseline set of calls that can be made from at the sequencer level as a command or from within the user code. The DL and DDL are usually just kept as linkable libraries, and the user code is compiled and linked prior to upload to the boat. This provides an efficient development environment for quickly prototyping code on the water if need be.

4. HAMMERHEAD STEREO SOFTWARE

The core stages and substages of the Hammerhead processing pipeline (referring to Figure 4) are as follows:

1. Image processing:
   (a) Stereo ranging: generate dense range images from the left/right stereo pairs
   (b) Plane finding: find the water surface and compute a stabilized camera pose
2. Mapping
   (a) Projection: project range data into a 2D grid map centered on the vessel
   (b) Filtering: perform spatial and temporal filtering on the map
   (c) Classification: compute a hazard probability for each map cell
3. Tracking
   (a) Detection: detect discrete objects (contacts) in the 2D grid map
   (b) Classification: assign a type to each contact
   (c) Tracking: fuse contacts into tracks and estimate speed and heading

The pipeline produces two outputs: a grid-based hazard map, suitable for use by static navigation behaviors, and a

discrete contact list, suitable for use by dynamic navigation behaviors. Some objects, such as boats, may appear in both. The pipeline also has two classification stages: the first assigns a hazard probability to each map cell, and the second assigns a type to each contact (e.g., boat, yacht, navigation marker, or false alarm). The classifiers are generated offline using a supervised learning algorithm with hand-labeled training data.

Two key complications arise in applying this algorithm pipeline to the maritime environment. First, the unusual arrangement of cameras on the Hammerhead sensor head (see Figure 5) reduces the effective resolution. Stereo camera pairs are normally constructed such that the image planes of the two cameras are approximately aligned; this allows us to create a pair of virtual rectified cameras that have very similar geometry (and thus similar field of view and resolution) to the real cameras. On Hammerhead, the virtual rectified cameras are rotated with respect to the real cameras, which leads to nonuniform resolution in the rectified images. Specifically, the rectified images are compressed in the forward direction (relative to the boat) and stretched at the sides. This forward compression is particularly troublesome, because it reduces the apparent size of objects and thus the effective detection range. Consequently, in order to maintain adequate resolution in the forward direction without sacrificing field of view at the sides, the rectified images must be oversized by approximately 20% relative to the raw images.

Second, the range data must be filtered to remove noise. The stereo correlator includes a number of internal checks, but some false correspondences inevitably make their way into the disparity image. In many applications, these errors can be removed using simple postprocessing techniques, segmenting the disparity image and removing segments that have a small numbers of pixels or subtend a very small area. In the maritime environment, however, this approach tends to remove both the water surface and the objects of interest (which are generally small). We therefore augment the standard postprocessing techniques with a simple multiscale filter: stereo ranging is performed independently at multiple scales (downsampled images), and inconsistent results are removed. This approach generates a better signal-to-noise ratio for on-water images and requires only a modest increase in total computation (stereo correlation scales as the cube of image size, so halving the image size reduces computation by a factor of eight).

Application of the multiscale filter produces a relatively clean map centered on the boat (example shown in Figure 7) that can be used for hazard detection. The map is georegistered with the GPS information from the INS, and static hazards are labeled based on the DNC. The size of the error ellipses indicates the uncertainty in the stereo range information, which grows with relative distance more in the aligned depth direction than the cross-track direction due to pixels of mixed range on the boundaries of the objects. The grid map is passed on to the tracking stage, which detects, classifies, and tracks discrete objects such as boats, channel markers, and buoys. The output of the tracking stage is a set of contacts describing the position, velocity, and type of sensed objects in the world. Further details of the stereo processing pipeline are available from the Office of Naval Research, Code 33.

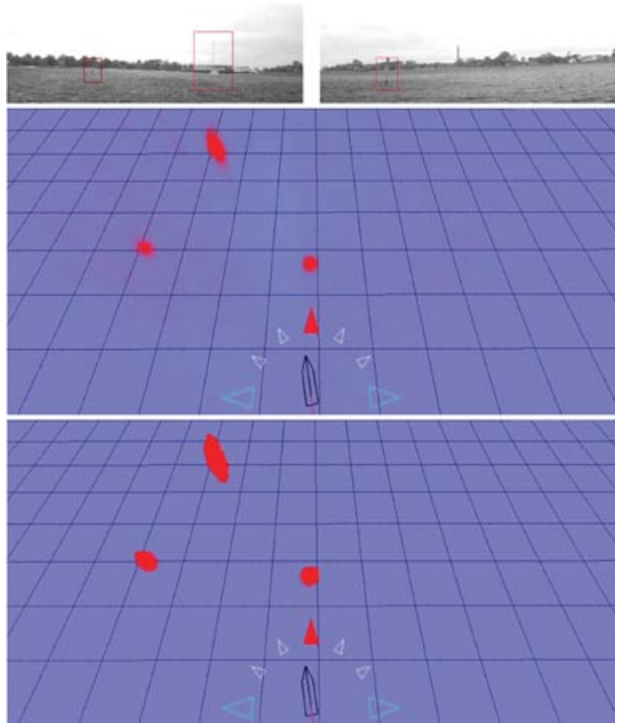

Figure 7. Mapping example from Hampton Reach channel entrance. (Top) Input images from the left-facing and right-facing camera pairs. Three objects (indicated on the images) are within sensor range: a channel marker due north, a second channel marker to the northwest, and a yacht to the north-northwest. The bounding boxes have been drawn manually for clarity. (Middle) Posterior mean elevation map (brighter shades of red denote larger elevations). The USV is shown at the bottom center of the map, heading slightly west of due north. (Bottom) Output of the map classifier, the obstacles marked in red. Note that the three objects are clearly resolved and the map contains no false alarms.

5. HAMMERHEAD EXPERIMENTAL STUDIES

The Hammerhead system was extensively tested over the past 3 years at Ft. Monroe, Virginia, in a number of different settings including riverine and blue water environments. The test range is shown in Figure 8. All runs were done with a safety officer onboard because the Norfolk channel

Figure 8. Test range used to exercise the Hammerhead stereo system. The James River zone used for the Trident Warrior 2009 fleet exercise was wide relative to the Hampton Channel. The blue water zone ranges from 3 to 50 m in depth.

and St. James River have heavy boat traffic. Ground truth for path was taken with a standard WAAS-enabled GPS receiver that has a position uncertainty of 3–5 m.

The Hammerhead system produces two distinct outputs: a grid-based hazard map and a list of discrete contacts. In this section, we evaluate the end-to-end performance for both kinds of outputs. Figure 9 depicts the four data sets used in this assessment, each of which captures a different scenario or environment:

a. The *crossing* data set is a slow traverse in open water while a target boat cuts back and forth ahead.
b. The *channel* data set is a fast traverse in open water with occasional channel markers.
c. The *following* data set is a slow traverse in open water while following a target boat.
d. The *river* data set is a slow traverse in a narrow, cluttered riverine environment.

The first three data sets in Figure 9 have been ground truthed to enable quantitative evaluation. This is a manual process in which a human operator draws a bounding box around each discrete object (such as boats or channel markers) in each image frame. The final data set is far too complicated to ground truth in this manner and is included to allow qualitative evaluation of mapping in complex environments.

5.1. Hazard Mapping

Quantitative evaluation of the hazard map is fraught with difficulties. First, the ground-truth data are expensive or impossible to obtain in all but the simplest scenes, and second, metrics of performance tend not be very meaningful. Consider, for example, a pair of maps that differ in only two cells: one of the maps correctly detects a channel marker in front of the vessel but misses a portion of the shoreline; the other detects the shoreline but completely misses the channel marker. These two maps may have the same detection and false-alarm rates (measured on a per-cell basis) but will result in radically different outcomes when used to plan safe paths. For these reasons, our evaluation of the Hammerhead hazard map is mostly qualitative.

Figure 10 shows some examples of the map generated for the *river* data set [Figure 9(d)]: the map is color-coded to indicate hazard probability. The map has been overlaid with data from an ENC that shows the surveyed position of the channel markers, shorelines, and occasional shore features. Note the overall correspondence between objects in the map and objects in the chart, as well as the absence of false alarms. Discrepancies in object location, where present, are due to uncertainty in the vessel position (GPS accuracy is approximately 3–5 m) or inconsistencies in the chart (which does not always reflect reality). On the other hand, many objects that are present in the chart are

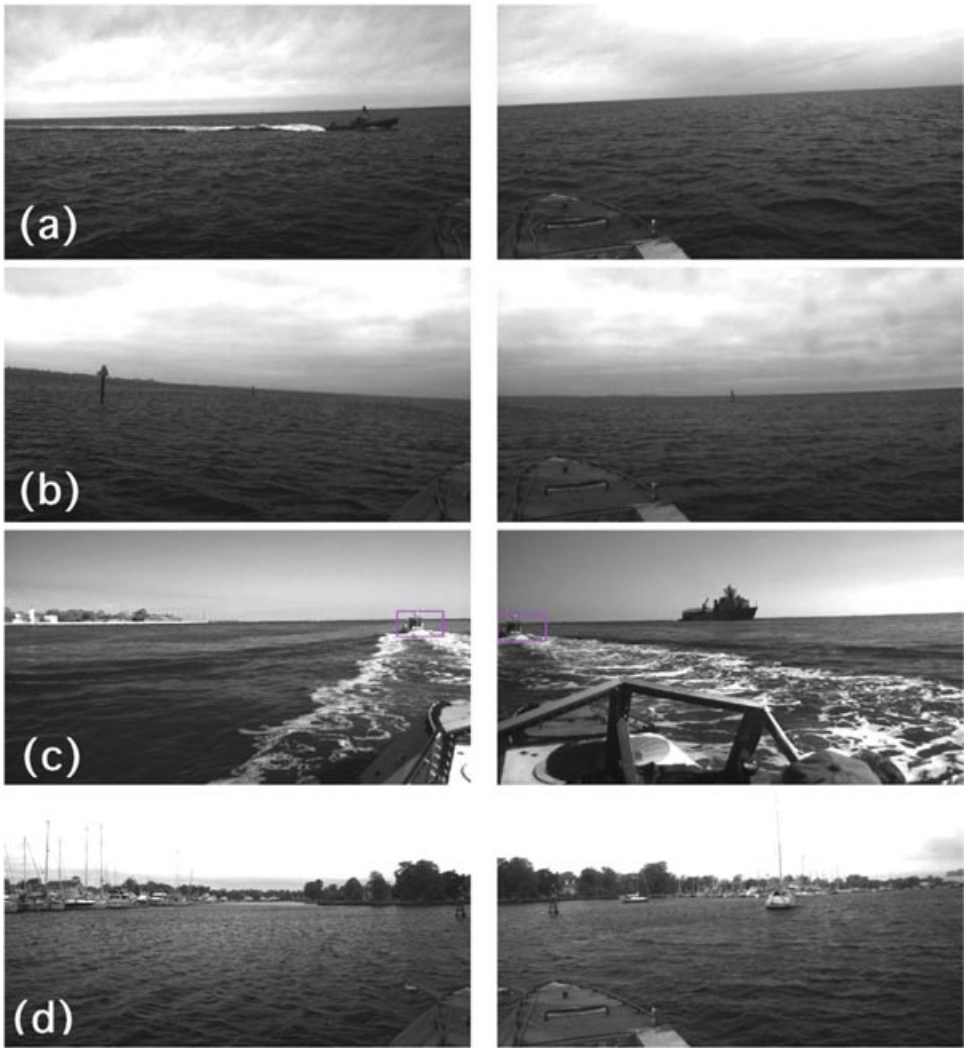

Figure 9. Sample stereo images from the four evaluation data sets. (a) *Crossing*: slow traverse on open water while a target boat repeatedly crosses our bow; (b) *channel*: fast traverse on open water with isolated channel markers; (c) *following*: slow traverse following a target boat that remains directly ahead; (d) *river*: slow traverse of a narrow, cluttered riverine environment.

missing from the map, because the objects are out of range, occluded, or simply not detected by the Hammerhead algorithm. Some areas of the shoreline are in the last category: the hazard detector is designed to find objects protruding above the waterline and therefore performs poorly on gently sloping beaches.

5.2. Contact Detection and Classification

Unlike the mapping stage, contact detection and classification can be evaluated quantitatively using the ground-truthed data sets. Following the approach of previous JPL work on vision-based pedestrian detection (Bajracharya, Moghaddam, Howard, Brennan, & Matthies, 2009), we use the detection probability (DP), defined as the total number of objects detected divided by the true number of objects as a metric.

The *crossing* and *channel* data sets [Figures 9(a) and 9(b)] were chosen for illustration because they contain the two most important object classes—boats and channel markers—at a variety of ranges. The analysis uses an 80/20 cross-validation scheme, in which 80% of the frames are randomly selected as the training set and the remaining 20% are held as test data. The averaged result of repeated application of this process gives a quantitative measurement of the algorithm. Two additional points should be noted. First, the system is trained using examples from both data sets but is tested using samples from either the

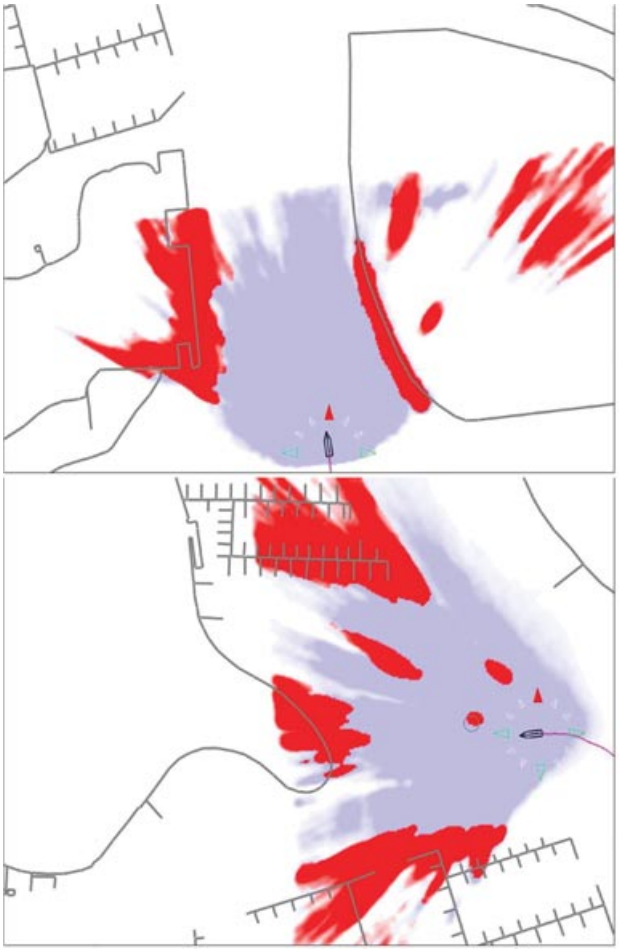

Figure 10. Example hazard maps from the *river* data set. Darker (red in color version) shades denote cells with higher hazard probabilities. Gray (blue in color version) indicates where there is stereo coverage generated by the Hammerhead system.

crossing or channel data sets. Because the first data set consists entirely of boats and the second entirely of channel markers, this allows us to assess the performance of these two object types individually. Second, the system is trained using examples from all ranges but tested using samples from a limited range interval to allow us to evaluate the effective range of the system. A comparison of the detection rates for the two data sets is shown in Figure 11 and is discussed further below.

The classifier was trained on 990 positive and 380 negative examples (drawn from both data sets) and evaluated on 260 samples drawn from the *crossing* data set. A 100% detection rate (DP) was not achieved at any range, and this effect is particularly pronounced for the long-range intervals (where DP is at most 70%). Digging into these missed detections, one finds that the fault lies not with the classifier, detector, or map but rather with original stereo range data. At longer ranges, the target boat subtends only a handful of pixels and the window-based stereo algorithm has difficulty detecting the boat reliably.

Next, the *channel* data set that consists entirely of channel markers was analyzed. The classifier is identical to that used for the *crossing* data set but is evaluated on 200 samples drawn solely from the *channel* data set. Compared with the boat, the channel markers clearly present a more challenging target. In practice, of course, we must operate the system with a single contact probability threshold. Using the optimal results from the crossing data set, we obtain a DP of 85%. At this detection rate, the tracker is able to bridge the missing frames, such that we maintain a reliable "lock" on channel markers.

Comparing these two data sets, the effective detection range for channel markers appears to be greater than that for boats, despite the fact that the markers are significantly smaller. We attribute this difference to the fact that the markers are stationary and the boats are in motion; this allows the map to accumulate more range points, which boosts the detection rate. The weak point for channel markers is the classifier, which has difficulty distinguishing between a relatively small upright feature and comparably sized false alarm.

As a final test, we have also evaluated the detector/classifier performance on the *following* data set (a target boat viewed from astern) using the model generated from the first two data sets. This result acts as a sanity check: it confirms that a classifier trained on one type of boat and viewpoint will generalize to a different type of boat and viewpoint. In 2,300 frames of the *following* data set [Figure 9(c)], with varying ranges, the Hammerhead system achieves a detection rate of 100% with no false positives.

5.3. Contact Tracking

Figure 12 shows the position and velocity estimates produced by the tracker for the first few minutes of the *following* data set [Figure 9(c)]. In this set, the USV is trailing a target boat more or less directly astern. The first set of plots shows the position estimates relative to the USV. We plot the position in the USV frame, such that the along-track axis is aligned with the bow of the USV and the cross-track axis is at right angles. Over the course of this run, the estimated along-track 3σ root-mean-square (RMS) uncertainty is 1.65 times that of the cross-track uncertainty. This is an expected property for systems based on stereo ranging, in which angular resolution (which determines cross-track uncertainty) is much better than range resolution (which determines along-track uncertainty).

The second set of plots shows the velocity estimates relative to the USV. The along-track 3σ RMS uncertainty is 1.25 times that of the cross-track uncertainty. Given that the speed is estimated from changes in position, this is an

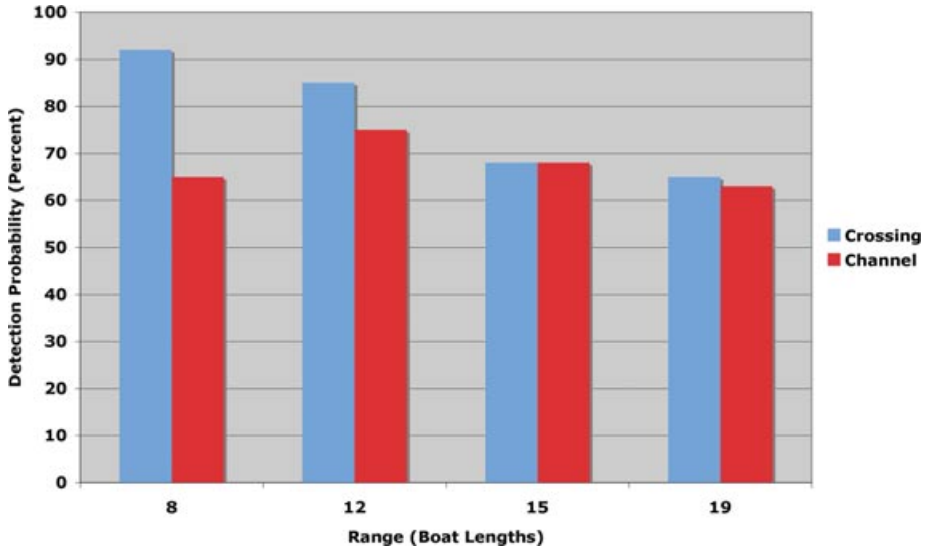

Figure 11. Comparison of the experimental results for the *crossing* and *channel* data sets. The range is given in boat lengths because this is the appropriate metric for safe navigation. See text for further discussion.

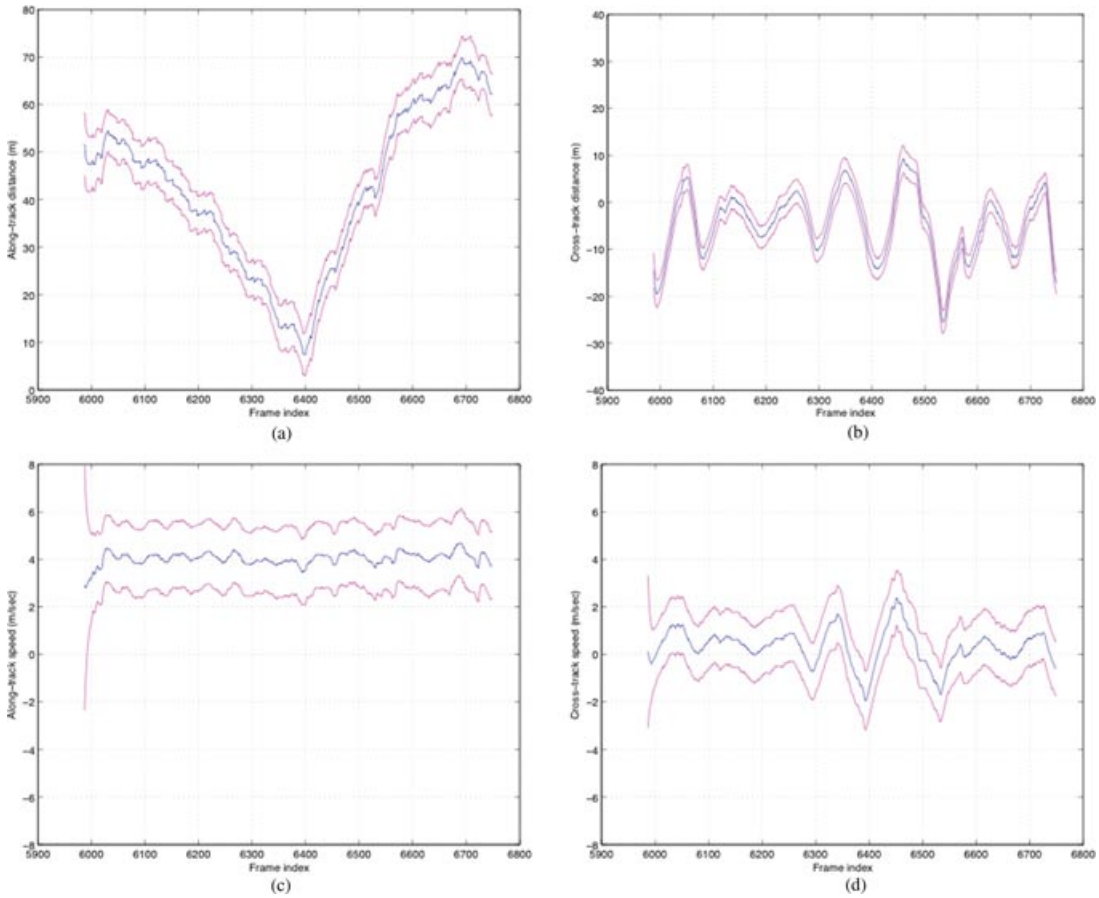

Figure 12. Tracking results for the first few minutes of the *following* data set. (a) Distance to the target boat, in meters. The upper and lower curves show the 3σ uncertainty bounds; (b) cross-track distance to the target boat, in meters; (c) along-track speed of the target boat, in meters per second; (d) cross-track speed to the target boat, in meters per second.

expected result. The estimated speed for the target boat is within 4% of that reported by the pilot's manual controls.

Ideally, we would like to evaluate the accuracy of the position and velocity estimates produced by the tracker by comparing them with ground-truth values. In practice, however, accurate ground-truth data are very difficult to obtain. For example, standard WAAS-enabled GPS receivers have a position uncertainty of 3–5 m, which is comparable to the tracking errors that we are trying to measure. Future experiments may use more precise differential GPS (DGPS) systems to improve the quality of the ground-truth data.

6. INTEGRATED TEST RESULTS

The Hammerhead stereo system is one component in a larger autonomy package for USVs. The hazard map and contact list are sent to the R4SA system, which implements the hazard avoidance and following behaviors. In this section, we briefly describe the implementation of these behaviors and present results from integrated field trials.

6.1. Local Hazard Avoidance

R4SA implements local hazard avoidance using a simple but effective arc-based planner. The algorithm operates on a regular grid that is centered on the vessel and aligned with the current heading. Hazard maps from the Hammerhead system are transformed into this grid, and cells containing obstacles (with high probability) are marked as nontraversable. A set of fixed arcs is tested for traversability against the grid, and the arc that is both traversable and closest to the goal is executed.

Arcs are tested extremely quickly through the use of a precomputed lookup table, as illustrated in Figure 13. For each grid cell, the table records the set of arcs that intersect that cell including the width of the boat. When hazards are projected into the grid, the algorithm can identify all arcs that intersect the cell and therefore all arcs that intersect the hazard. Thus, a single pass over the grid is sufficient to determine which arcs are nontraversable. This lookup table is computed once at program initialization, works for arbitrary arc-generating functions, and can include factors such as the boat size and safety margins.

Note that the maritime environment has a number of unique features that make planning both easier and harder than it is in ground-based robotics (for example). On the one hand, the environment is relatively uncluttered, such that simple algorithms such as the one described above will handle most situations. On the other hand, the environment is intrinsically dynamic; oceans and rivers have currents, so paths must account for the motion of the water as well as the motion of the boat. In the present system, this is done by assuming an upper bound on the speed of the current and expanding the safety margins accordingly.

The integrated Hammerhead/R4SA hazard avoidance behavior has been tested over a cumulative distance of about 1,500 nautical miles and demonstrated on a number of occasions, including the Trident Warrior naval exercises in June 2009. Figure 14 shows two segments of the trajectory executed by the CMV USV while it was running under autonomous control. The vessel passes under the James River Bridge (near Newport News, Virginia), adjusting its path to pass through the center between the bridge pylons, and then passing through. Note that the turns are very smooth, indicating that the Hammerhead system was able to detect the hazards in good time and that R4SA was able find suitable obstacle-free paths.

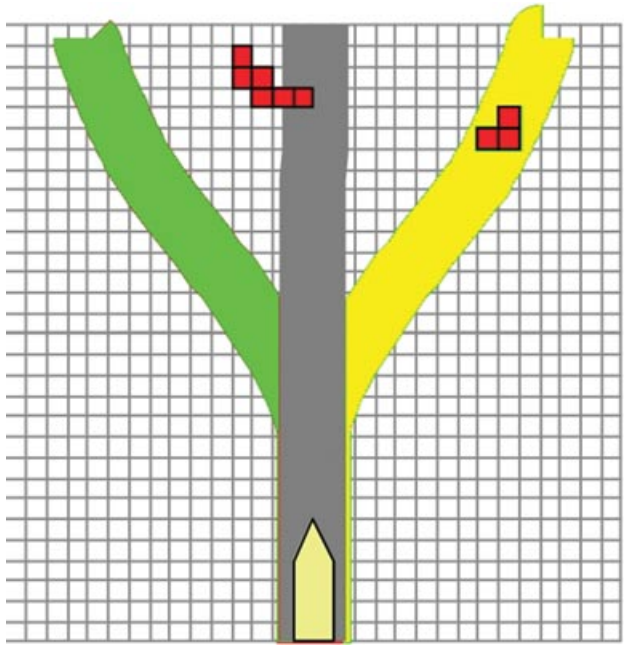

Figure 13. R4SA hazard avoidance algorithm. This example shows three arcs, two of which are blocked by hazards (center and right paths). Each grid cell also records the set of arcs that intersect that cell, so a single pass through the hazard map is sufficient to determine which arcs are blocked.

6.2. Target Following

R4SA implements a following behavior that allows the USV to approach a target vessel and trail it at a set distance. The algorithm is a variant of the closest point of approach (CPA) algorithm described in Nicholson (2008) and makes use of the contact position and velocity estimates provided by the Hammerhead system. Given these data, R4SA computes a goal location that is fixed with respect to the target (e.g., directly astern) but is moving along a vector parallel with the world frame. R4SA servos the USV to this goal location using a proportional–integral–derivative (PID) control loop that also matches the speed and heading of the target vessel. Gain scheduling is used to ensure safe operation; that is, R4SA uses different PID control loops, depending

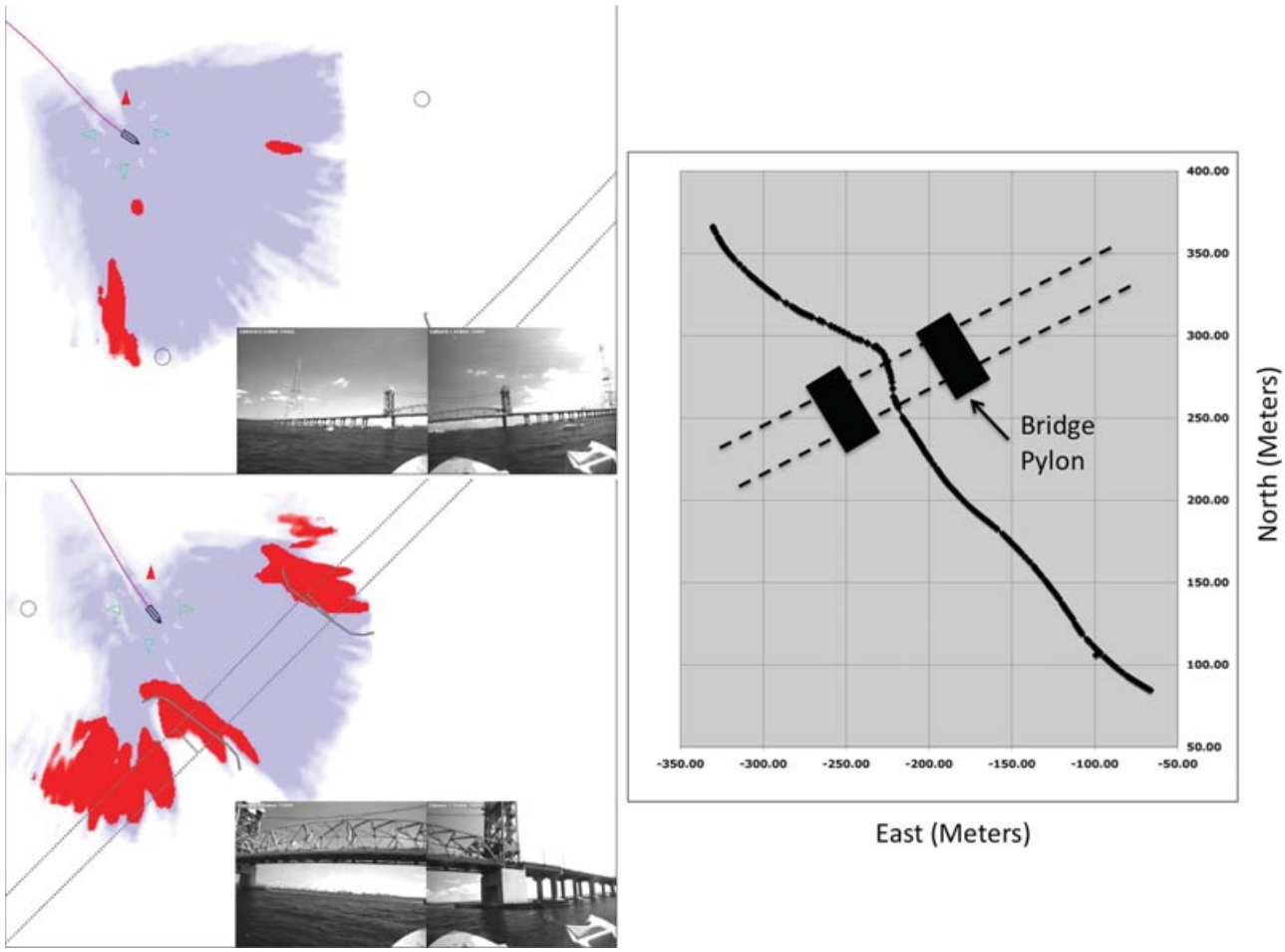

Figure 14. Integrated hazard detection and avoidance during the Trident Warrior 2009 naval exercises in Norfolk, Virginia. The CMV USV passes under the James River Bridge near Newport News, avoiding a boat (upper left) and the bridge pylons (lower left). Right: The ground truth GPS plot, where the course correction to pass midway between the pylons can be clearly seen.

on the proximity to the target vessel. At short ranges, the controller is very aggressive, commanding large deviations in the rudder and throttle to ensure that the USV does not collide with the target vessel.

Figure 15 shows one example of the following behavior from an integrated field trial in November 2009. The USV first approaches and then shadows the target vessel, following it for several minutes across the harbor. Figures 15(b) and 15(c) show the relative distances (along track and cross track); after the initial acquisition, R4SA is able to rapidly servo to the moving goal and maintains a fixed position with respect to the target. The change in range at around the 30-s mark is due to an operator command to R4SA, requesting an increase in the boat separation. Not shown in these figures are a number of other boats that are detected and tracked by Hammerhead but ignored by R4SA (which remains locked on the target).

The target vessel in these trials is neither cooperative nor noncooperative; it simply holds a more-or-less fixed course and speed. Under these conditions, we observe that the integrated following behavior is very stable so long as the target vessel remains within the Hammerhead field of view. We have not yet tested with noncooperative targets, which we expect to be significantly more challenging.

7. CONCLUSIONS AND FURTHER WORK

The Hammerhead vision system has undergone a series of successful on-water tests in a number of different environments including cluttered river channels, harbors, and littoral zones. The testing was carried out under a fairly wide range of conditions, including light to moderate sea states, light rain, and low sun angles. The system has been

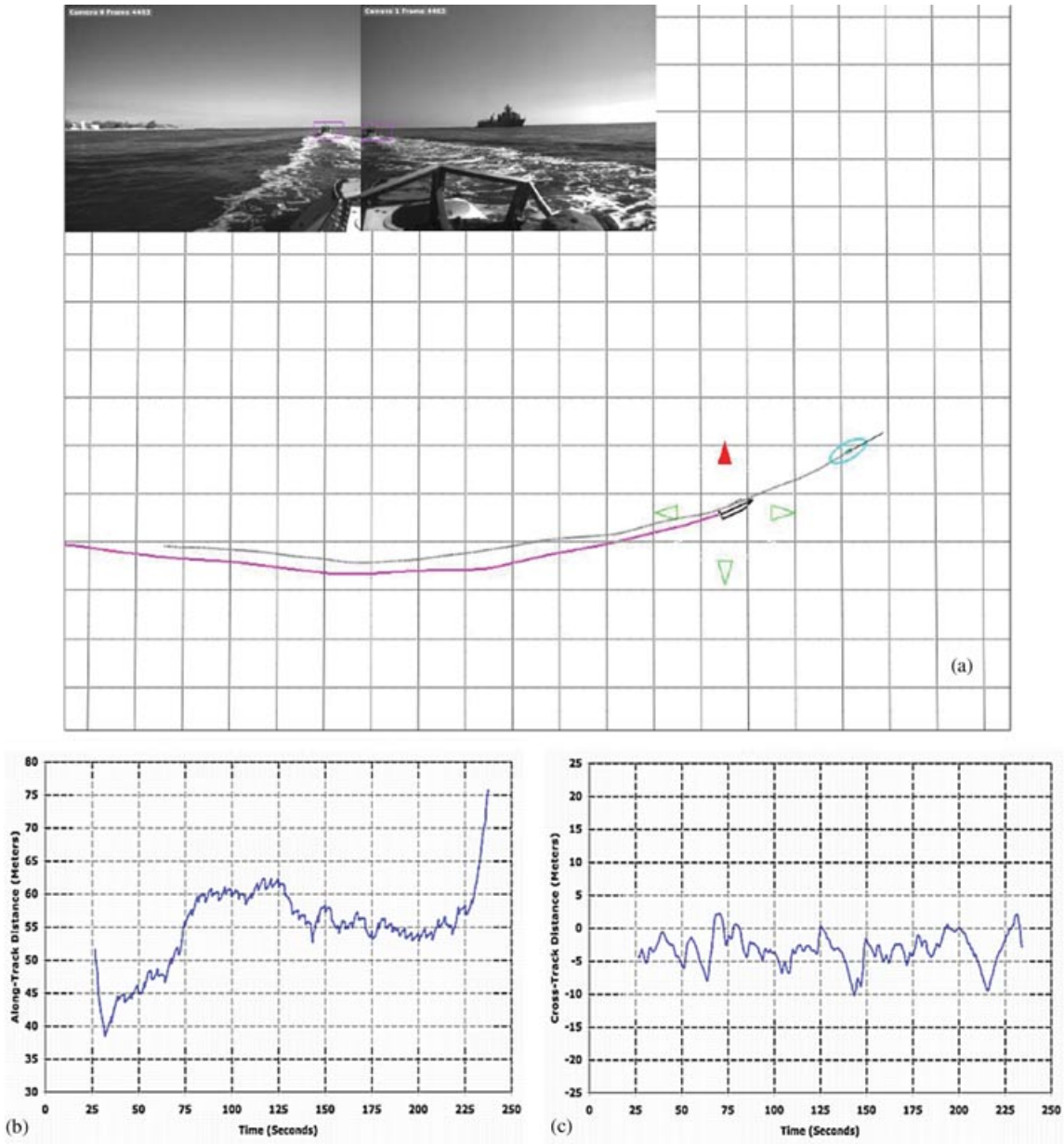

Figure 15. Integrated following tests with the HTF USV. Plot of the USV and target trajectories for a short segment. (a) Along-track distance to the target boat as a function of frame number (total duration of 4 min); (b) cross-track distance to the target boat as a function of frame number.

run in autonomous mode for a total of more than 1,500 nautical miles over the past 3 years. The accuracy of the sea surface identification degrades with higher sea states (higher waves) due to increasing occlusion. Robust results have been achieved up to sea state 2 (waves of 35–70 cm in height). A commercial Rain-Off spray was applied to the camera windows, and the water rolled off in light rain up to 6 mm/h with some spotting and loss of stereo information but not enough to degrade the navigation performance. The sensors are mounted on the arch about 3 m over the water, so splashing occurs only at high speed and/or in high sea states. We are currently testing in higher sea states and at higher speeds. Polarization filters on the camera lenses cut the glare appreciably, and stereo performance for navigation was robust from 1 h after dawn to 1 h before dusk while heading directly into the sun.

Two outstanding issues remain, however. The first is lighting: we are currently studying the use of long-wave infrared (IR) cameras to enable day/night operation. In principle, the software is agnostic to the details of the camera sensor, and the current processing pipeline (with some retuning and retraining) is expected to work just as well with IR cameras as it does with visible light cameras. The second issue is range: the current system has an effective look-ahead distance that is relatively close for higher speed navigation (>35 knots) on water. We are therefore considering the use of higher resolution cameras, coupled with new algorithms and/or dedicated hardware for stereo ranging. Each doubling of the camera resolution also doubles the effective look-ahead distance but increases the stereo processing requirements by a factor of eight.

On the integrated Hammerhead/R4SA system, we have only just begun testing of vision-based navigation behaviors. One issue that remains is the correct way to handle static versus dynamic objects. This distinction is particularly important for obeying maritime rules of the road (COLREGS). Testing is currently underway on a dynamic path planning method for COLREGS compliance. The current system makes no distinction between the two types of objects, and neither the grid-based hazard map nor the discrete contact list is an entirely appropriate representation. For example, the contact list describes the velocity of each object but is not guaranteed to be complete (it does not contain extended obstacles such as shorelines, for example). In contrast, the hazard map is complete but does not contain velocity estimates. Finding the best representation to enable full and safe autonomy in a dynamic maritime environment remains an open question.

ACKNOWLEDGMENTS

The research described in this paper was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. Funding for this work was provided by the Office of Naval Research, Code 33 (Contract N00014-09-IP-2-0008), and Spatial Integrated Systems, Inc. (NASA Space Act Agreement, Contract NMO716027). The authors would like to thank the reviewers for their comments that greatly enhanced the readability of the manuscript. Finally, we wish to thank Curtis Padgett, Mike Garrett, Lee Magnone, Harry Balian, David Zhu, and Eric Kulczycki for their valuable contributions to this project.

REFERENCES

Angelova, A., Matthies, L., Helmick, D., & Perona, P. (2007). Learning and prediction of slip from visual information. Journal of Field Robotics, 24(3), 205–231.

Bajracharya, M., Moghaddam, B., Howard, A., Brennan, S., & Matthies, L. H. (2009). A fast stereo-based system for detecting and tracking pedestrians from a moving vehicle. International Journal of Robotics Research, 28(11–12), 1466–1485.

Bandyopadhyay, T., Sarcione, L., & Hover, F. (2009, July). A simple reactive obstacle avoidance algorithm and its application in Singapore Harbour. In Proceedings of the 7th International Conference on Field and Service Robotics, Cambridge, MA (to appear).

Elkins, E., Sellers, D., & Monarch, W. R. (2010). The Autonomous Maritime Navigation (AMN) Project: Field tests, autonomous and cooperative behaviors, data fusion, sensors, and vehicles. Journal of Field Robotics, Special Issue on State of the Art in Maritime Autonomous Surface and Underwater Vehicles, Part 1, 27(6), 790–818.

Goldberg, S. B., Maimone, M. W., & Matthies, L. (2002, March). Stereo vision and rover navigation software for planetary exploration. In 2002 IEEE Aerospace Conference, Big Sky, MT (pp. 2024–2036).

Hansen, E., Huntsberger, T., & Elkins, L. (2006, April). Autonomous maritime navigation: Developing autonomy skill sets for USVs. In D. W. Gage (Ed.), Proceedings of SPIE, Unmanned Systems Technology VIII, Orlando, FL (vol. 6230, pp. 221–232).

Helmick, D., A. Angelova, A., & L. Matthies, L. (2009). Terrain adaptive navigation for planetary rovers. Journal of Field Robotics, 26(4), 391–410.

Huntsberger, T., Aghazarian, H., Castano, A., Woodward, G., Padgett, C., Gaines, D., & Buzzell, C. (2008, June). Intelligent autonomy for unmanned sea surface and underwater vehicles. In AUVSI Unmanned Systems North America, San Diego, CA.

Iagnemma, K., Shibly, H., & Dubowsky, S. (2002, May). Online traction parameter estimation for planetary rovers. In IEEE International Conference on Robotics and Automation, Washington, DC.

Larson, J., Bruch, M., Halterman, R., Rogers, J., & Webster, R. (2007, August). Advances in autonomous obstacle avoidance for unmanned surface vehicles. In AUVSI Unmanned Systems North America, Washington, DC.

Larson, J., Ebken, J., & Bruch, M. (2006, March). Autonomous navigation and obstacle avoidance for unmanned surface vehicles. In SPIE Unmanned Systems Technology VIII, Defense Security Symposium, Orlando, FL (vol. 6230).

Matthies, L. (1989). Dynamic stereo vision. Ph.D. thesis, CMU-CS-89-195, Department of Computer Science, Carnegie Mellon University.

Nicholson, M. (2008). Maneuvering board manual (PUB 217). National Geospatial Intelligence Agency. Annapolis, MD: Lighthouse Press.

Ruiz, A. R. J., & Granja, F. S. (2009). A short-range ship navigation system based on LADAR imaging and target tracking for improved safety and efficiency. IEEE Transactions on Intelligent Transportation Systems, 10(1), 186–197.

Snyder, F. D., Morris, D. D., Haley, P. H., Collins, R., & Okerholm, A. M. (2004, March). Autonomous river

navigation. In D. W. Gage (Ed.), Proceedings of SPIE, Mobile Robots XVII, Orlando, FL (vol. 5609, pp. 221– 232).

Talukder, A., Manduchi, R., Rankin, A., Owens, K., Matthies, L., Castano, A., & Hogg, R. (2002, October). Autonomous terrain characterization and modeling for dynamic control of unmanned vehicles. In IEEE/RSJ International Conference on Intelligent Robots and Systems, Lausanne, Switzerland.

Wolf, M. T., Assad, C., Kuwata, Y., Howard, A., Aghazarian, H., Zhu, D., Lu, T., Trebi-Ollennu, A., & Huntsberger, T. (2010). 360-degree visual detection and target tracking on an autonomous surface vehicle. Journal of Field Robotics, Special Issue on State of the Art in Maritime Autonomous Surface and Underwater Vehicles, Part 1, 27(6), 819–833.

Zhang, W., Zhuang, P., Elkins, L., Simon, R., Gore, D., Cogar, J., Hildebrand, K., & Crawford, S. (2009, March). A stereo camera system for autonomous maritime navigation (AMN) vehicles. In SPIE Conference on Unmanned Systems Technology, Orlando, FL (vol. 7332).