Passive perception system for day/night autonomous off-road navigation

Arturo L. Rankin*[a], Charles F. Bergh[a], Steven B. Goldberg[b], Paolo Bellutta[a], Andres Huertas[a], Larry H. Matthies[a]
[a]Jet Propulsion Laboratory, 4800 Oak Grove Drive, Pasadena, CA, USA 91109
[b]Indelible Systems, 8955 Quartz Ave, Northridge, CA, USA 91311

ABSTRACT

Passive perception of terrain features is a vital requirement for military related unmanned autonomous vehicle operations, especially under electromagnetic signature management conditions. As a member of Team Raptor, the Jet Propulsion Laboratory developed a self-contained passive perception system under the DARPA funded PerceptOR program. An environmentally protected forward-looking sensor head was designed and fabricated in-house to straddle an off-the-shelf pan-tilt unit. The sensor head contained three color cameras for multi-baseline daytime stereo ranging, a pair of cooled mid-wave infrared cameras for nighttime stereo ranging, and supporting electronics to synchronize captured imagery. Narrow-baseline stereo provided improved range data density in cluttered terrain, while wide-baseline stereo provided more accurate ranging for operation at higher speeds in relatively open areas. The passive perception system processed stereo images and outputted over a local area network terrain maps containing elevation, terrain type, and detected hazards. A novel software architecture was designed and implemented to distribute the data processing on a 533MHz quad 7410 PowerPC single board computer under the VxWorks real-time operating system. This architecture, which is general enough to operate on N processors, has been subsequently tested on Pentium-based processors under Windows and Linux, and a Sparc based-processor under Unix. The passive perception system was operated during FY04 PerceptOR program evaluations at Fort A. P. Hill, Virginia, and Yuma Proving Ground, Arizona. This paper discusses the Team Raptor passive perception system hardware and software design, implementation, and performance, and describes a road map to faster and improved passive perception.

Keywords: Passive perception, stereo vision, autonomous navigation, sensor head, infrared, parallel processing, FPGA

1. INTRODUCTION

Perception for Off-Road Robots (PerceptOR) was one of six key supporting technology programs of the Defense Advanced Research Projects Agency (DARPA)/Army Future Combat System (FCS) program. PerceptOR was a three year program which ran from March 2001 until March 2004. It was designed to develop prototype perception approaches to advance the state-of-the-art for off-road obstacle detection, and to enable higher levels of autonomous mobility needed for FCS operations. Four contractor awards were made at the beginning of the first year, followed by competitive down selects at the beginning of the second and third year. Science Applications International Corporation (SAIC) received one of the awards under DARPA agreement MDA972-01-9-0015. The SAIC led team, called Team Raptor, was composed of SAIC, the Machine Vision Group (MVG) at the Jet Propulsion Laboratory (JPL), Carnegie Mellon University (CMU), Applied Perception Incorporated (API), and Visteon Corporation. SAIC automated an off-the-shelf (OTS) all-terrain vehicle (ATV), implemented navigation sensors and a computing and communications backbone, provided regulated power distribution and a charging system, implemented a world model and path planning, and developed an operator control unit (OCU). API developed an active perception system (with support from CMU and Visteon), which included, a Riegl two-axis scanning ladar, multiple SICK single-axis scanning ladars, and multiple automotive radar sensors. JPL-MVG developed a passive perception system that included a variety of passive imaging sensors to enable day and night perception under reduced magnetic signature conditions[1]. The Team Raptor robotic vehicle was named “Raptor”.

* Arturo.L.Rankin@jpl.nasa.gov; phone (818) 354-9269; fax (818) 393-4085

The passive and active perception systems provided the world model with sensor-based terrain maps. Raptor could be configured or commanded to operate using solely passive perception, active passive perception, or both. The passive perception system was used during extensive data collections at Ft. Polk, LA, Chatfield State Park, CO, Buffalo Creek, CO, and on undeveloped properties near SAIC, Littleton, CO. In addition, the passive perception system was operated during autonomous navigation evaluations on sponsor selected courses at Ft. A.P. Hill, VA, and Yuma Proving Ground (YPG), AZ. This paper describes Team Raptor's passive perception hardware and software design, implementation, and performance, and a road map to faster improved passive perception.

1. PASSIVE SENSORS

The PerceptOR program was divided into three phases. Phases 1, 2, and 3 lasted approximately 8 months, 12 months, and 16 months, respectively. During Phase 1, an OTS Honda Rubicon ATV was automated and outfitted with sensors and computing hardware, and rudimentary detection and avoidance of discrete obstacles was demonstrated during the daytime. Due to the long lead time of infrared cameras, a pair of Amber Radiance 1 mid-wave infrared (MWIR) cameras was borrowed from JPL for this demonstration. Since the Phase 1 passive sensor package was composed primarily of non-PerceptOR funded sensors, it will not be discussed in this paper. The next two sections discuss the Phase 2 and Phase 3 passive sensor head design and implementation.

2.1 Phase 2 Sensor Head Design and Implementation

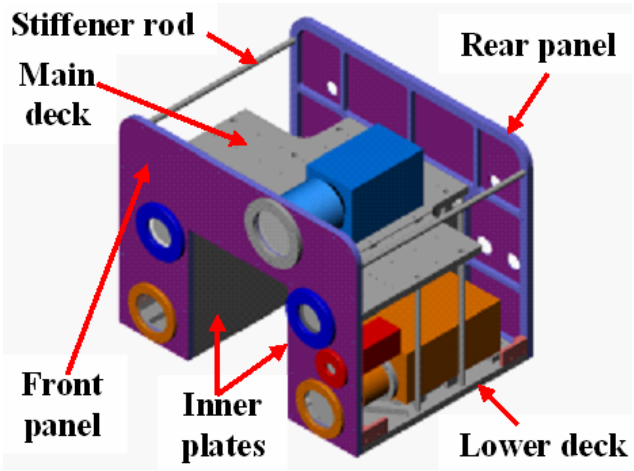
During Phase 2, JPL designed and fabricated an environmentally protected forward-looking passive sensor head and purchased two OTS camera enclosures for a stereo pair of rear-looking color cameras. This section will discuss only the design and implementation of the forward-looking sensor head. The forward-looking sensor head was designed (using SolidWorks 3D modeling software) to straddle a Vinten Incorporated HS102PE industrial size pan-tilt unit (PTU). It is composed of five primary structural components: a lower optical deck to house two infrared cameras, a main optical deck to house all other cameras and peripheral electronics, two vertical inner plates to connect the main and lower optical decks and provide a mounting interface to the PTU, 2.1mm thick aluminum removable front and rear panels, and a 1.3mm thick aluminum U-shaped removable access cover. Six 9.5mm diameter stiffener rods were added to the sensor head to prevent flexure of the lower deck relative to the main deck, and to prevent flexure of the front face panel. Figure 1a illustrates the location of the primary structural components.

The main optical deck is the central component of the sensor head structure. It was fabricated from 12.7 mm thick 6061-T6 aluminum tooling stock. Ribs were cut into the bottom of the main deck in order to achieve the desired stiffness and lower the system weight. The resulting shell thickness was 2.2mm. Pass through features were added to the rear portion of the deck to allow cabling and hand access between the two decks. A series of #6-32 tapped holes were added (on the center lines of the support ribs) to form an optical bench. The rear panel contained bulkhead connectors to interface power and computing hardware. The front panel contained O-ring sealed, environmental windows to provide optical filtering for each camera and to seal the sensor head from the environment. The removable access cover sat on a neoprene rubber gasket and was held in place by four quick-lock case fasteners.

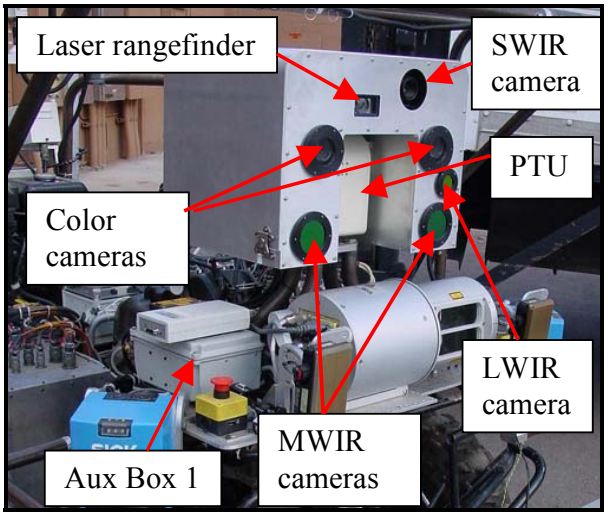
A model of each Phase 2 sensor was included in the computer aided design (CAD). The Phase 2 forward-looking sensor head design and implementation (including sensors) is illustrated in Figures 1a and 1b. The sensor head included:

- a pair of Hitachi HV-C20 3CCD color cameras with a baseline of 30 cm,
- a pair of Cincinnati Electronics NC256 cooled, InSb based, 3.6-5 μm MWIR cameras with a baseline of 33.5 cm,
- an Indigo Systems Alpha uncooled, microbolometer based, 7.5-13.5 μm long-wave infrared (LWIR) camera,
- a Sensors Unlimited SU320M-1.7RT 0.9-1.7 μm short-wave infrared (SWIR) camera,
- and a Leica Disto Pro non-scanning, 635 nm single point red laser rangefinder.

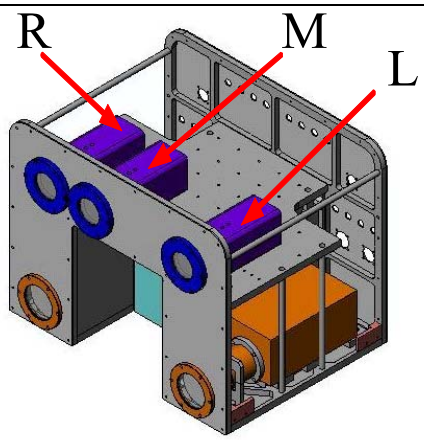
The MWIR cameras were mounted to the lower deck, the color cameras and LWIR camera were mounted to the underside of the main deck, and the remainder of the sensors was mounted on the top of the main deck. Finite element



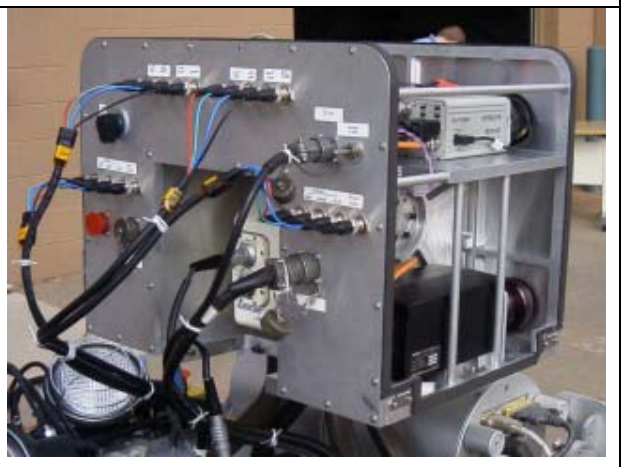
(a) 3D CAD model of the Phase 2 forward-looking passive perception sensor head (minus the laser rangefinder).



(b) The Phase 2 forward-looking passive perception sensor head with the service cover installed.



(c) 3D CAD model of the Phase 3 forward-looking sensor head. The left (L), middle (M), and right (R) color cameras are labeled.



(d) Side view of the Phase 3 forward-looking passive perception sensor head with the service cover removed.

Figure 1: The Phase 2 and Phase 3 forward-looking passive perception sensor head design and implementation.

analysis (FEA) was performed on the design (using NEiNastran FEA software) to evaluate the stresses each sensor would see during worst case operation. Total displacement was limited to less than 10% of a pixel. In addition, each sensor was positioned to be accessed easily by hand and removed without affecting the calibration of the other sensors.

Vinten originally designed the PTU to mount a single television camera on one side of the unit. The advertised payload capacity of the unit is 15.75 kg. The angular range, maximum velocity, and maximum acceleration is +/- 179°, 60°/s, and 180°/s^2, respectively, for both axes. The accuracy and repeatability are 0.016° and 0.01°, respectively. Two modifications were made to the OTS unit. The first, made by Vinten, was to include a mounting bracket on both sides of the unit. The second modification, made by Team Raptor, was the installation of the two U.S. Digital AD4-B-S external quadrature-to-RS232 adapters to tap into the pan and tilt encoders for software position feedback.

The red laser rangefinder supported preliminary experiments in sensing water depth (i.e., fordability). The color and MWIR cameras supported passive stereo ranging and obstacle detection; the MWIR camera pair provided for operation at night and in atmospheric obscurants. The color, LWIR, and SWIR cameras also supported terrain classification; the

LWIR especially for night operation and the SWIR especially for water detection during the day. Although hardware was implemented in the forward-looking sensor head to support a Duncan Tech MS2100 multispectral camera (on loan to the PerceptOR program from JPL for data collections), it was not used during the program.

The composite video output signal from the LWIR camera was used to trigger the other cameras. (Two Burst Electronics MicroDA-4 video distribution amplifiers were used to distribute the sync signal.) Figure 2 illustrates synchronized imagery from each type of camera within the Phase 2 forward-looking sensor head. All camera video signals, control signals, and power were bundled into waterproof cables with locking bayonet connectors to run between the forward-looking sensor head and the Mobility Behavior Control Unit (MBCU). Two weatherized auxiliary electronic boxes (Aux Box) containing passive sensor and PTU support components (such as the quadrature-to-RS232 adapters and the video distribution amplifiers) were designed and installed on Raptor.

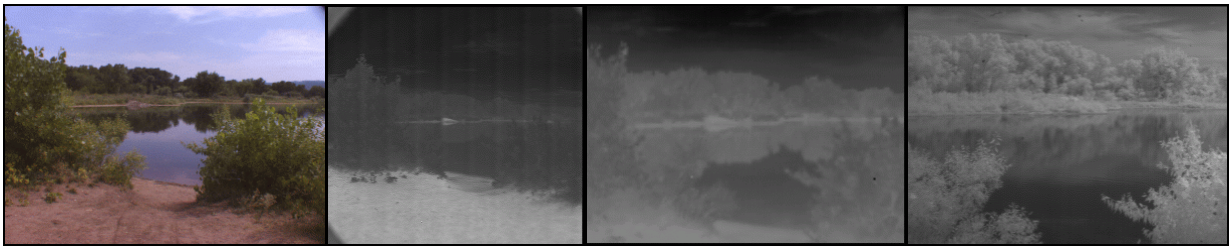


Figure 2: Images from the forward-looking sensor head acquired at Chatfield State Park, CO, August 1, 2002, 3:25pm. From left to right, RGB color, MWIR, LWIR, and SWIR images.

2.1 Phase 3 Sensor Head Design and Implementation

The use of a single stereo pair with a fixed baseline required trading off range resolution for stereo matching performance on tall grass. While cost prohibited adding a third NC256 MWIR camera for stereo ranging at night, the addition of a third color camera to support stereo ranging during the daytime with three baselines was inexpensive. During Phase 3, the front passive sensor head was modified to accommodate a third HV-C20 color camera. Our primary motivation was to provide narrower baseline options to enable better stereo performance in challenging terrain, enable comparison of performance with different baselines, and allow the use of a larger baseline at higher speed.

During Phase 2, the uncooled LWIR camera suffered from poor contrast and high levels of structured noise in cool weather. The SWIR camera and the Leica visible laser rangefinder performed adequately. The work in Phase 3, however, required a shift in focus to deal with navigating through cluttered terrain, such as the forests at Ft. A.P. Hill. The LWIR, SWIR, and laser sensors were removed from the front passive sensor head in order to focus on multiple baseline stereo. The Phase 3 forward-looking sensor head design is illustrated in Figure 1c. A rear view of the forward-looking sensor head (with the service cover removed) is shown in Figure 1d. The final forward-looking sensor head mass and outer dimensions are 16.55 kg (including sensors) and 47.6 cm wide x 35.6 cm deep x 36.8 cm tall. The center of mass was at (0.05, -0.82, 1.7) cm from the intersection of the pan and tilt axes. The slight offset in rotation about tilt axis (Y) is intentional. This balances the force exerted by the cabling on the rear of the sensor head. Peripheral support electronics were used to trim the balance on the sensor head, making ballast unnecessary.

Table 1. Primary specifications of the passive sensors in the Phase 3 forward-looking sensor head.

Sensor Description	Manufacturer Model	Size (HxWxL) cm	Weight (g)	Power	Effective Field of View	Sensor Resolution
3.6-5 µm cooled MWIR camera	Cincinnati Electronics NC256	7.1x7.1x29.5	1845.8	15VDC, 45W startup, 16W steady state	67.5°x62.4°	256x256
3 CCD color (RGB) camera	Hitachi HV-C20	6.5x6.5x17.3	716.5	10.5-17VDC 4.5W	55.8°x42.7°	640x480

Table 1 shows the primary specifications for the Phase 3 forward-looking passive sensors. The HV-C20 cameras are high quality 3CCD cameras that accept a standard C-mount lens, and capture interlaced images. As the two fields are captured at different times, the even field is dropped and only the odd field is stored and processed. The HV-C20 can automatically control exposure, gain, and lens iris for best image quality. But to reduce the effects of motion blur, we elected to maintain a fixed exposure time. The NC256 MWIR cameras use a Stirling-cycle cooled Focal Plane Array (FPA) of 256x256 pixels with a F2.0 cold shield. These cameras have a standard DIOP bayonet lens mount. A Janos Technology Incorporated 7mm ASIO lens is installed on each camera. The lens, which does not have any iris adjustment, has a fixed F2.0 aperture. SAIC designed and fabricated clamps to anchor each ASIO lens to the lower optical deck. After startup, the MWIR imagery is unusable for stereo vision for up to 6 minutes while the FPA cools to approximately 80° Kelvin.

During Phase 3, a video sync generator (Video Accessory Corporation VB/BBG-3) was added to the forward-looking passive sensor head to sync each camera. As shown in Figure 3, a star distribution pattern was used. In additional, an add-on Aux Box was fabricated and installed on Raptor to allow software resetting of PTU. Due to physical constraints, the PTU could not achieve the advertised range of motion on Raptor. The length of the PTU power and data cables limits the pan range of motion to approximately ±45º about the forward-looking position. Straight-through data and power connectors on the rear of the PTU, and the close proximity of the Riegl laser scanner, limits the tilt range of motion to approximately ±15º about the forward-looking position. The combination of the camera field of view (FOV) and PTU range of motion enables a field of regard (FOR) of 145ºx72º by the forward-looking color cameras and 157ºx152º by the forward-looking infrared (FLIR) cameras. Panning across the range of motion (–45º to +45º) at the maximum pan speed (60º/s) takes approximately 1.5 seconds.

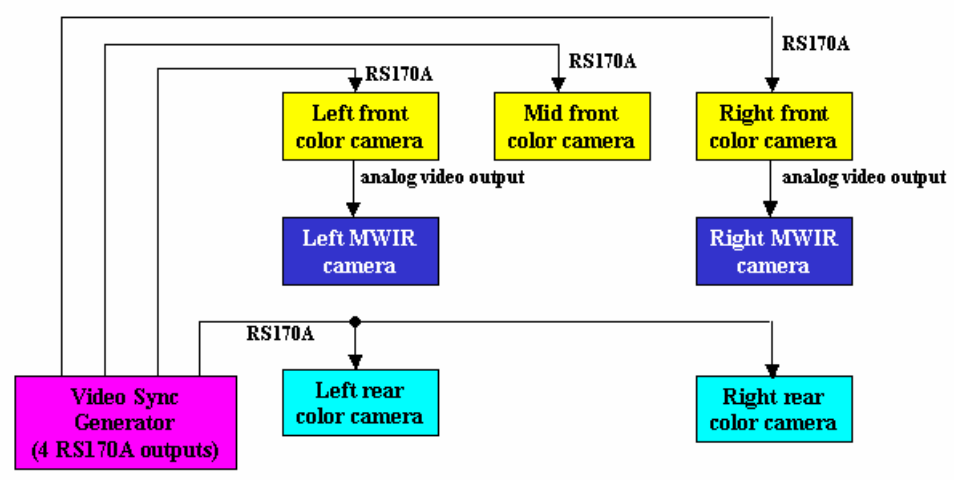
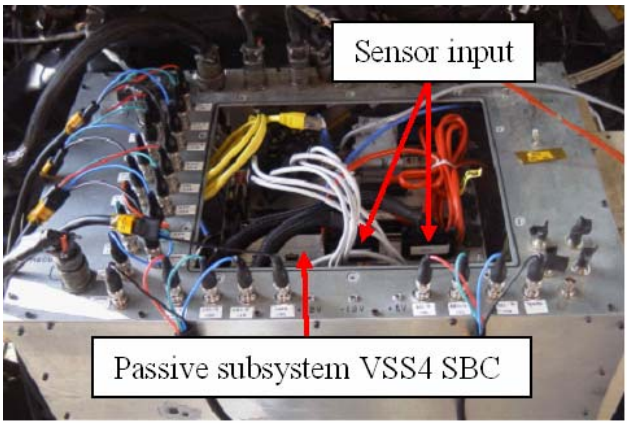


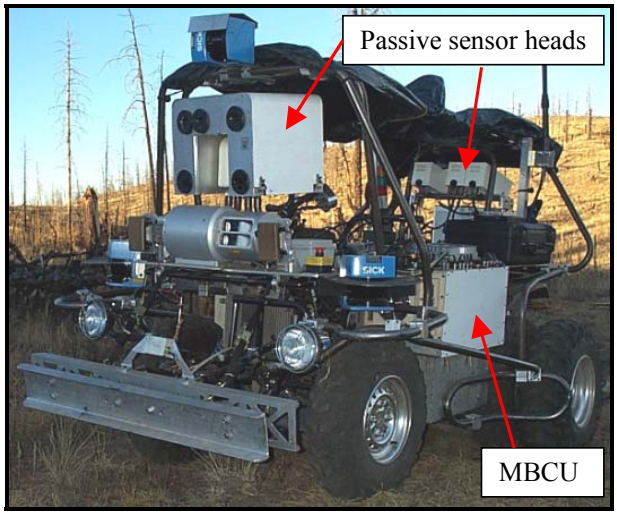
Figure 3: The genlock port of each color camera is connected to an output line from the video sync generator. The left and right MWIR cameras are triggered through the composite video output of the left and right color cameras, respectively.

3. COMPUTING HARDWARE

The passive perception system computing hardware is housed within the MBCU, an aluminum enclosure custom-built by SAIC that provides thermal management, shock-damping, and environmental (dust and moisture) protection. The MBCU contains a custom-built VME computer card cage that houses a host SAIC mobility behavior processor board, a passive perception system processor board, a 36GB disk drive for the host board, a 120GB disk drive for the passive perception processor board, a real-time kinematic GPS receiver, and an ethernet hub. The disk drives, manufactured by Red Rock Technologies, were composed of ATA hard disks mounted to 6U VMEbus boards with an ultra wide SCSI interface. A picture of the MBCU (with its service cover removed) is shown in Figure 4a. Figure 4b shows the location of the MBCU on the Raptor vehicle in its final configuration.



(a) The MBCU with its service cover removed.



(b) The Raptor vehicle in its final configuration.

Figure 4: The SAIC MBCU housed the passive perception system computing hardware.

During Phases 1 and 2, the passive perception processor board was the Synergy MicroSystems VGM5, PowerPC 7400 Single Board Computer (SBC), in the 6U form factor. The VGM5, which contains a single processor, operated under the VxWorks 5.4 operating system. During Phase 3, the passive perception processor board was upgraded to a 533MHz Synergy MicroSystems VSS4, which is a 6U VMEbus SBC containing four PowerPC 7410 microprocessors. Each processor is provided with 2 megabytes (MB) L2 backside cache. The board provides 512MB DRAM, Ultra Wide SCSI controller, 100/10BaseT Ethernet controller, four RS232 serial ports, one PMC slot, 64MB of user Flash, 128 kilobytes (KB) NVRAM and a real time clock/calendar with four digits for the year. The passive perception system processor board (VSS4) and the mobility behavior processor board are connected to the MBCU ethernet hub and communicate over the network, rather than over the VME bus. The passive perception processor board booted off the host SCSI disk during Phase 2 and the dedicated passive perception SCSI disk during Phase 3. A startup script is specified in the VSS4 boot parameters, enabling the passive perception system to automatically start when the MBCU box is powered on.

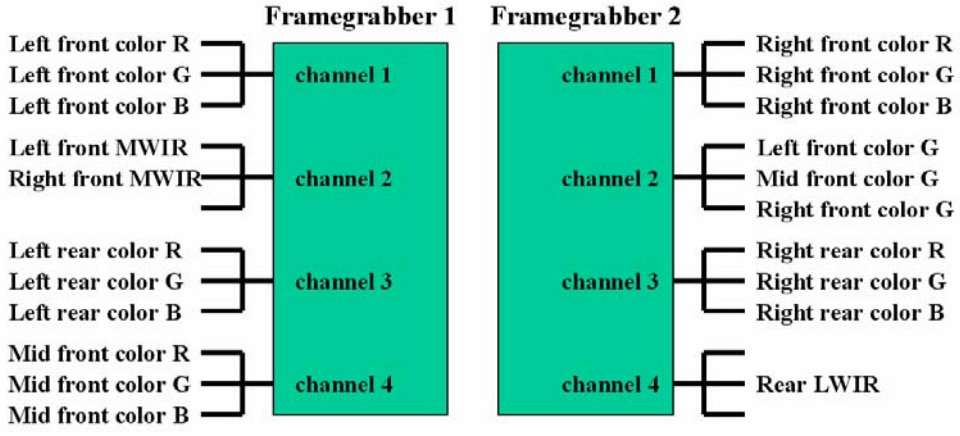


Figure 5: The Phase 3 sensor interface to the framegrabbers.

The VSS4 supports multiprocessing with high-bandwidth DRAM memory shared by each CPU, and a private mailbox for each CPU. A Synergy MicroSystems PU32 PCI Mezzanine Card (PMC) serial port module is installed in the VSS4's single on-board PMC slot. Five serial ports are used: two RS232 ports to configure the NC256 MWIR cameras (ports 4 and 5), two RS323 ports to read pan (port 8) and tilt (port 7) encoder values, and one RS422 port used to issue

pan and tilt commands to the internal PTU controller (port 9). A Synergy Microsystems PEX3 PMC carrier board, which provides three additional PMC slots, is installed on the VSS4. Two Active Silicon Snapper PMC-24 analog framegrabbers are installed on the PEX3. The PMC-24 digitizes input imagery at 8 bits/pixel/color at a resolution of 640x480 image resolution.

Figure 5 shows the final Phase 3 framegrabber wiring diagram. Each PMC-24 contains a quad mux providing four RGB channels; a selected channel on each framegrabber can be acquired simultaneously. Here, we will refer to the simultaneous acquisition of imagery from channel X of framegrabber 1 and channel Y of framegrabber 2 simply as X-Y. Forward-looking wide, mid, and narrow-baseline stereo ranging could be performed during the daytime by selecting 1-2, 1-2, and 4-2, respectively. The green bands of the color images were used to perform daytime stereo ranging. Daytime rear-looking stereo ranging could be performed by selecting 3-3. Nighttime forward-looking stereo ranging could be performed by selecting 2-Y, where Y could be any channel on framegrabber 2. Selecting 2-2 could enable forward-looking daytime (all three baselines) and nighttime stereo ranging.

The selection of forward-looking wide, mid, or narrow-baseline stereo enabled front color classification and front color or monochrome imagery for teleoperation. The selection of rear-looking stereo enabled rear color classification and rear color or monochrome imagery for teleoperation. The selection of forward-looking MWIR stereo enabled front infrared imagery for thermal classification and nighttime teleoperation. In addition, it could enable front color classification and front color or monochrome imagery for teleoperation (2-1), or rear color classification and rear color or monochrome imagery for teleoperation (2-3). The selection of X-4 enabled rear infrared imagery for nighttime teleoperation.

4. SOFTWARE ARCHITECTURE

The passive perception software is composed of three primary modules: a data acquisition (DAQ) module, a run-time module, and an external interface module. The DAQ module contains interfaces to all the passive perception cameras and the PTU. The run-time module contains all the algorithmic software required to process imagery and generate passive perception terrain maps. The external interface module passes passive perception terrain maps (generated by the algorithmic software) to an SAIC world map module called TMAP, where terrain maps from the active and passive perception systems are fused. In addition, the external interface module processes a limited number of commands from the OCU (routed through TMAP), such as sensor and gaze control selection. During Phase 2, the DAQ, run-time, and external interface software modules ran on the same processor. At the end of Phase 2, interfaces existed that would allow us to extend this same three-part architecture onto multiple processors. Running the passive perception software on multiple processors increases the rate at which passive terrain maps are sent to TMAP.

The VSS4 quad CPU board is usually used with an operating system capable of Symmetrical Multi Processing (SMP), a computer architecture in which the operating system is tasked with inter-processor communications and load balancing. However, the VxWorks 5.4 operating system does not have SMP support. Rather, VxWorks required us to run four copies of the operating system in four segments of the VSS4's 512MB of memory. We did not want to special case our run-time system around this rather unique multi-processor architecture, so we developed an inter-processor architecture that uses TCP/IP sockets to pass messages. This allowed us to develop and test our distributed run-time system on four separate computers without having to rely on shared memory. Once we had debugged this method, we special cased the message passing for shared memory to pass pointers to large data structures rather than passing data structures over the internal IP network. Having successfully built a distributed run-time system, we partitioned the work amongst the four VSS4 processors X, Y, Z and W (as named by Synergy). Since processor X is directly tied to all available input-output (IO) interfaces (frame grabbers, serial ports, and the external network), and any external IO done by Y, Z or W must pass through X, processor X was the logical place for the DAQ module. This also made processor X the logical place for a multicast image server that supports OCU teleoperation, and the endpoint for sending passive terrain maps to TMAP via a Neutral Machine Language (NML) communications buffer. Processor X also runs the server side of the multi-processor architecture, serving images to processors Y, Z and W. This architecture is commonly referred to as farming.

Figure 6 illustrates the multi-processor software architecture. As suggested by the yellow boxes, each of the three coprocessors runs the same run-time module executable as processor X. Different master processor and coprocessor

configuration files direct the coprocessors to act as clients rather than servers. In addition, having different master and coprocessor configuration files allows different software components to be enabled. This reduced the overall complexity by requiring only a single build of the run-time module executable. Software components can be turned on or off before startup in the configuration files, or after startup, by telneting into each processor individually and running a utility called SetVar. Compartmentalizing the hardware specific parts of the passive perception software into the DAQ module has allowed the run-time source code to be completely platform independent. Compiling the run-time source code under Linux, Windows, and Mac OS X has allowed us to acquire data sets and simulate run-time results offline on multiple platforms. On the Raptor vehicle, separate DAQ, run-time, and external interface executables are compiled for VxWorks.

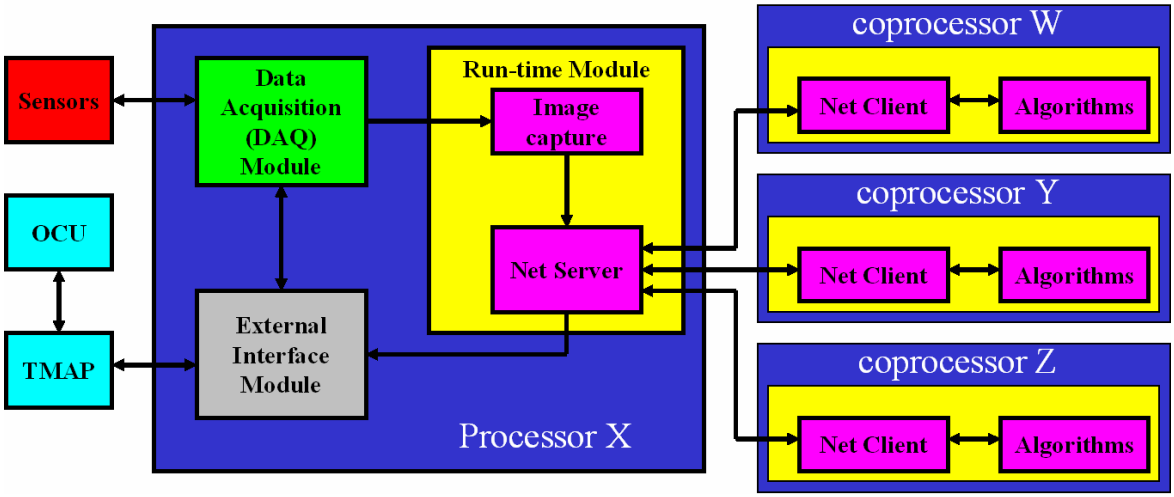


Figure 6: The passive perception system software architecture on one master processor (X) and three coprocessors (W, Y, and Z).

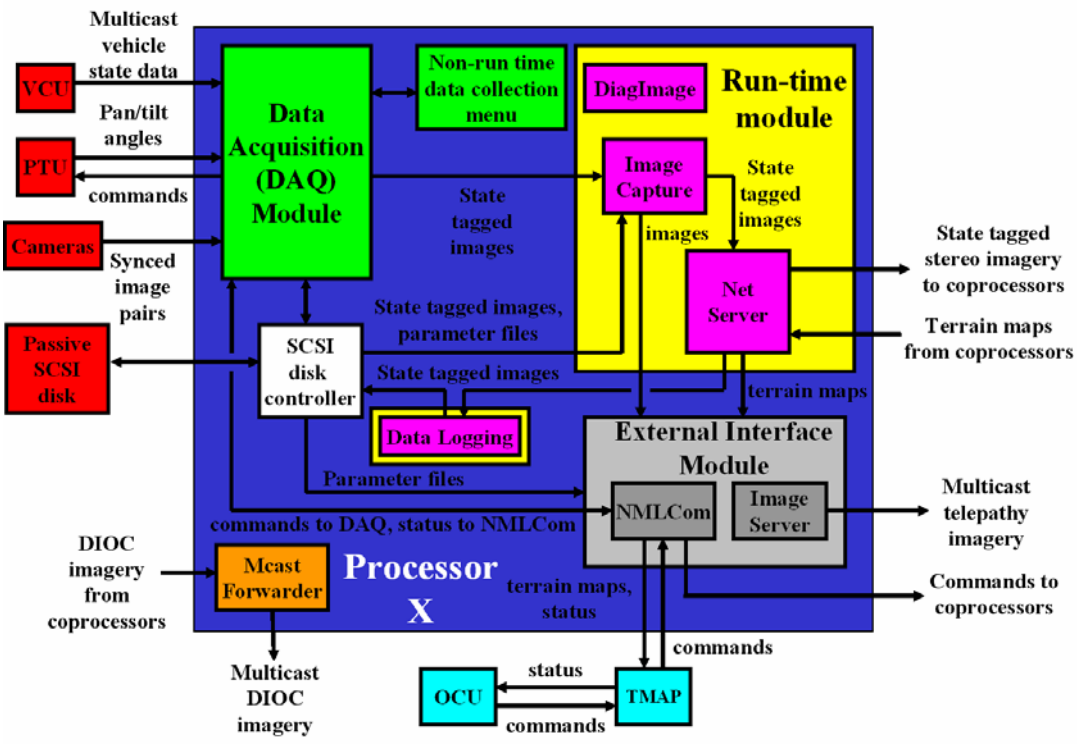


Figure 7: A more detailed view of the software the runs on processor X.

Figure 7 shows a more detailed diagram of the software components that are typically configured to run on processor X, and the flow of the data between them. The red boxes represent the hardware components that the passive perception system obtains data from. The SCSI disk controller (shown in white) is provided by the VxWorks operating system. All software components running on processor X have access to the dedicated passive perception SCSI disk via the SCSI disk controller.

4.1 Data acquisition module

The DAQ module acquires synchronized imagery from pairs of calibrated cameras, the current vehicle state, and the current PTU state. The vehicle's state is obtained by capturing a multicast packet broadcast by the Vehicle Control Unit (VCU). The PTU state is obtained by reading the two serial ports connected to the encoder quadrature-to-RS232 adapters. The DAQ module records the current vehicle and PTU state data in the header of imagery acquired from the cameras. The green boxes in Figure 7 represent the software components that are in the DAQ module executable. The DAQ module can be used to perform a non run-time data collection at a resolution of either 640x240 or 320x240. A text menu can be requested from the processor X command line that allows a user to select cameras, a PTU angle, image resolution, monochrome or color format (if available), a frame rate, the total number of frames to save, and a location to save the data. The DAQ module delivers the state tagged image pairs to the image capture component of the run-time module at 15Hz. When running in simulation mode, the image capture component retrieves state tagged image pairs from the dedicated passive perception SCSI disk (instead of DAQ).

Gaze control is performed by the DAQ software module. As the vehicle speed increases, command steer angles need to be limited to avoid vehicle tip over. Therefore, the horizontal field of regard necessary for path verification decreases with increasing vehicle speed. Two primary gaze behaviors were implemented. In the first, the gaze control algorithm commands the PTU to pan the sensor head about the current steer angle with a swath angle that is inversely proportional to the vehicle's speed. In the second, the gaze control algorithm commands the PTU to pan about the PTU home position, again with a swath angle that is inversely proportional to the vehicle's speed. The current steer angle and the PTU home position can be viewed as a reference angle. The PTU tilt angle is maintained at a constant value.

4.2 Run-time module

The run-time module includes all software components that generate results in real-time. The run-time software processes the pairs of synchronized imagery and generates passive perception terrain maps that contain elevation, obstacle, and classification layers. The coprocessors are configured to run the algorithm components of the run-time module. Network server and client components are included in the run-time module to serve and retrieve data. Processor X is configured to run the network server and the coprocessors are configured to run the network client. The image capture component forwards state tagged image pairs to the network server component for delivery to the coprocessors. Whenever the network server receives the most recent image pair from the image capture component, it cycles between the coprocessors. When it finds a coprocessor network client that just delivered a passive terrain map, it sends that coprocessor network client the new image pair. If a new image pair arrives at the network server before the previous images are served, the previous images are discarded. While the run-time algorithm components on that coprocessor processes the image pair, the network server retrieves the most recent image pair from the image capture component and continues the distribution process.

The image capture component also forwards images requested by the OCU operator to the image server component within the external interface module. These images are broadcast via multicast packets to any process subscribed to receive them. The data logging component logs all state tagged imagery sent to the coprocessors by the network server. This allowed us to reproduce problems observed during autonomous navigation in a simulation environment. The passive system could be configured to log data all the time to a hardware circular buffer of 1000 frames (166 seconds of data collected at 6 Hz). The OCU operator can turn passive perception data logging on and command the archiving of the circular buffer.

Figure 8 shows the common algorithm components in the run-time module that can be run on the coprocessors. Only the components used during Phase 3 will be discussed here. The camera component reads CAHVOR camera model

files from the SCSI disk at startup. It generates rectification tables and vehicle state aligned CAHV camera models. The decimation, rectification, and grayscale component receives a raw image pair from the net client component. The input image has a resolution of 640x240 images. First, the input images are decimated to 320x240. If either of the input images are RGB color images, they are converted to grayscale. The decimated grayscale and color images are then rectified using the rectification lookup tables. The down-sampled rectified grayscale images are passed to the pre-filter component. If the input images are RGB color images and the color classifier component is turned on, the color classifier uses the left down-sampled rectified color image to generate a rectified color classification image. If the input images are MWIR images and the thermal classifier component is turned on, the thermal classifier uses the left down-sampled rectified infrared image to generate a rectified thermal classification image. The contours component uses the down-sampled rectified grayscale intensity image to generate a contour image that contains positive and negative contours.

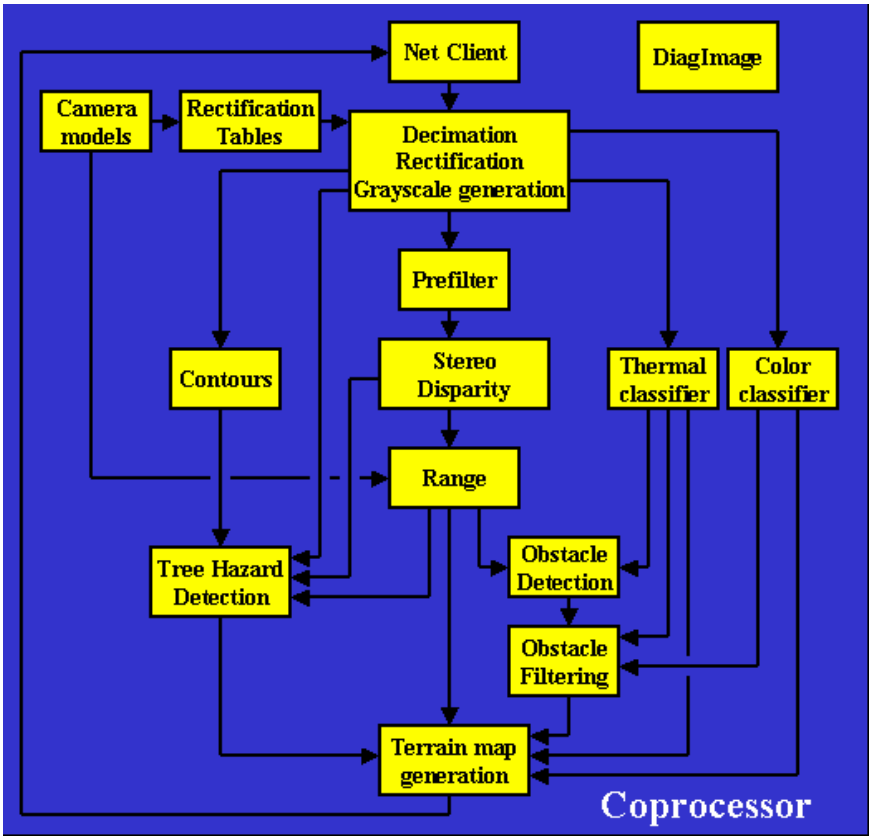


Figure 8: Passive perception algorithm components in the run-time module commonly enabled on the coprocessors.

The pre-filter component generates a normalized image by subtracting the local background from a rectified grayscale image. The stereo disparity component uses a pair of normalized images to generate disparity images and quality metrics by performing sum of absolute differences (SAD) correlation. It also applies a gradient sensitive region filter. If the density of the disparity image is above a parameter file threshold, a 3D floating-point stereo range image is generated using the disparity image and CAHV camera models.

The tree hazard component uses a down-sampled rectified grayscale intensity image, a contours image, a disparity image, and a range image to generate a tree hazard image. Trees are labeled as a severe, moderate, mild, or no hazard. The obstacle detection component contains a geometry based positive obstacle detector, a geometry based negative obstacle detector, and a thermal signature based negative obstacle detector. The geometry based obstacle detectors search each column of a range image, thresholding heights, slopes, and gaps, and output obstacle images. If the input imagery is from the MWIR cameras and the thermal signature based negative obstacle detector is turned on, the

thermally detected negative obstacles are added to the negative obstacle image. The obstacle filtering component performs 3D region filtering on the positive obstacle image and the negative obstacle image. If obstacle pruning is turned on, a classification image is used to determine the percent vegetation composition of each region. Small positive obstacles that are predominantly vegetation are pruned from the positive obstacle image.

The terrain map generation component uses a range image, a color or thermal classification image, a filtered positive obstacle image, and a filtered negative obstacle image, and a tree hazard image to produce a 2.5D, 30 meter compressed terrain map. Each 20cm x 20cm cell in the compressed terrain map contains the maximum ground cover elevation, a terrain classification type, an obstacle type, and confidence measures. High density, excessive slope, and low-overhang obstacle analysis is performed in the terrain map generation component.

Any diagnostic image written to the DiagImage component from any processor can be selected (from the processor command prompt) to be written to a file or the vehicle network. An OpenGL diagnostic display program (called glRemdisp) can be run on a remote computer (under Linux or Windows) to view the diagnostic images being written to the network. The terrain map generation component not only constructs, but also decodes compressed terrain maps, converting them to 150x150 pixel color-coded images, and sending them to the DiagImage component. Thus, passive terrain maps from any stereo baseline currently in use can be viewed remotely during run-time with glRemdisp. This is an extremely useful way to visualize a true representation of all the passive terrain maps being sent to TMAP by the external interface module.

4.3 External interface module

The external interface module contains two primary components: NMLCom provides a set of NML communication buffers with TMAP, and ImageServer multicasts low latency, high frame rate, low bandwidth telepathy imagery to subscribed clients, such as the OCU. When the network server component receives passive terrain maps from the coprocessors, the maps are forwarded to the NMLCom component of the external interface module for delivery to TMAP. NMLCom also sends a status packet to TMAP, approximately once per second, for delivery to the OCU. After startup, the first status packet is not sent until the first terrain map has been generated. Thus, the first status packet signals the OCU operator the passive system is running and ready to accept commands. The status packet reports selected sensors and stereo baseline, the current gaze control and obstacle detection sensitivity modes, sensor and PTU failures, the obstacle detectors currently enabled, and if the passive perception system is in live or simulation mode. If a terrain map has not been sent to TMAP within the last 2 seconds, the failure bits for the sensors currently selected are set. NMLCom also monitors the TMAP passive command buffer for commands forwarded to the passive system from the OCU. The commands the passive system can receive through the command buffer includes sensor and baseline selection for stereo processing, sensor selection for telepathy imagery, gaze control selection, requested obstacle detection sensitivity mode, enable or disable passive terrain map output to TMAP, data logging commands, reboot the passive system, and reset the PTU. Sensor and gaze control selection commands are forwarded to the DAQ module by the NMLCom component. Requests to change the obstacle detection sensitivity mode were forwarded to the coprocessors by the NMLCom component. The sensitivity modes are discussed in section 5.

4.4 Multicast forwarder module

In support of a requirement to display real-time diagnostic data in an analysis cell, JPL implemented an image server that packages a disparity image, intensity image, filtered obstacle image, and classification image into a single image format called DIOC. A DIOC image server can be enabled on each coprocessor. The intensity image data field is broken into two parts: an 8 bit data field and a 16 bit data field. If the input intensity image is grayscale, the 8 bit data field is populated and the 16 bit data field is eliminated. If the input intensity image is RGB color, the green data is assigned to the 8 bit data field and the red and blue data are assigned to the 16 bit data field. The 16 bit data field is at the end of the DIOC image. One can tell if a DIOC image has color intensity data by the size of the image. DIOC images contain vehicle and PTU state data in the header. In addition, the CAHVOR camera models are recorded in the image header. DIOC images are gzip compressed prior to broadcast. The coprocessor configuration file is used to enable or disable the DIOC image server. When enabled, the coprocessors multicast their DIOC images on the internal network. A multicast forwarder (running on the master processor) receives DIOC multicast images from the

coprocessors and forwards them onto the vehicle network. The multicast forwarder module is compiled as a stand-alone executable. Multicast and logged DIOC images can be read and displayed by an OpenGL diagnostic display program called 3rdEye. 3rdEye unpacks the contents of a DIOC image and regenerates range data (using the same stereo ranging algorithm on Raptor) from the disparity data and the camera models. The image intensity, classification, and obstacle pixels can be overlaid on the 3D range data. To reduce wireless network congestion, 3rdEye was used sparingly during run-time in Phase 3. It has been used extensively in simulation to visualize the 3D range data and to understand how new algorithm modifications affect range data.

5. PASSIVE PERCEPTION CAPABILITIES AND RESULTS

Our color classification, positive obstacle detection, and geometry based negative obstacle detection algorithms have been described in [2]. During Phase 2, the passive perception areas of focus included detection of water from SWIR imagery, dual-band infrared terrain classification, and thermal signature based negative obstacle detection. At large incidence angles, SWIR image intensity is dominated by reflected light. Water detection with a UGV mounted passive SWIR sensor is likely to be ineffectual beyond a couple of meters[3]. The Indigo Alpha LWIR camera had unacceptable levels of fixed pattern and other noise, making a robust dual-band MWIR/LWIR terrain classification difficult. The Phase 2 dual-band classifier was implemented with the standard minimum distance classifier using the Mahalanobis distance metric. A sample dual-band MWIR/LWIR terrain classification result is shown in Figure 9. Simulation and experimental results indicated that at times during the diurnal cycle, sandy soil and vegetation classes overlap in dual-band color space (using broad band MWIR and LWIR channels). Simulations of a hypothetical sensor with bands centered at 4 and 11μm showed the ability to effectively discriminate soils, rock, and vegetation.

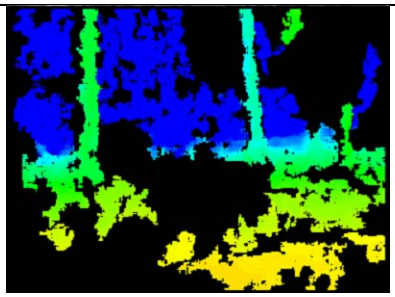
(a) MWIR image (b) LWIR image (c) Six class result

Figure 9: Sample nighttime (8:00pm) dual-band MWIR/LWIR terrain classification results of the scene in Figure 2.

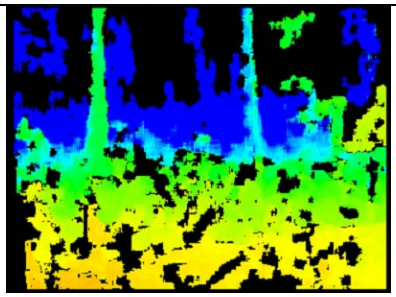
The work in Phase 3 required a shift in focus to deal with navigating through cluttered terrain. The Phase 3 areas of focus included multiple daytime stereo baselines, tree traversability analysis, combined thermal/geometry negative obstacle detection, and enhancements to a layered 2.5D terrain map representation. The Phase 3 passive perception system was operated during program evaluations at Fort A. P. Hill and Yuma Proving Ground. Sample Phase 3 results are shown in Figure 10. Figures 10b and 10c show false color wide and narrow-baseline range images for the forested scene in Figure 10a. (The pixels with colors close to red are closer to the sensors and the pixels with colors close to blue are far from the sensors.) The narrow-baseline range image is denser than the wide baseline range image, allowing the detection of the full extend of the fallen tree. However, the wide-baseline range resolution, as expressed in Equation 1, is five times greater than the narrow-baseline range resolution, where *IVFOV* is the vertical angular resolution of a pixel and *k* is a pixel precision factor. As a result, obstacles in narrow-baseline terrain maps tend to be more exaggerated in size than in wide-baseline terrain maps, particularly at further ranges.

$$\sigma_{Range} = \frac{Range^2 \cdot k \cdot IVFOV}{Baseline} \tag{1}$$

(a) Intensity image from Ft. A.P. Hill, VA

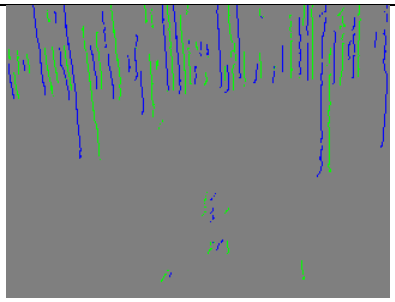


(b) 320x240 wide-baseline range image for the scene in (a)

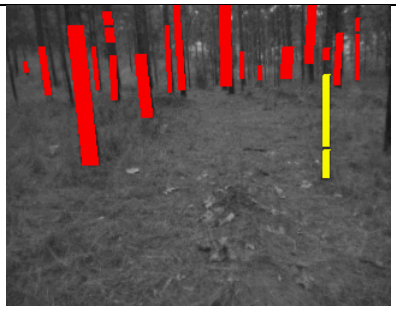


(c) 320x240 narrow-baseline range image for the scene in (a)

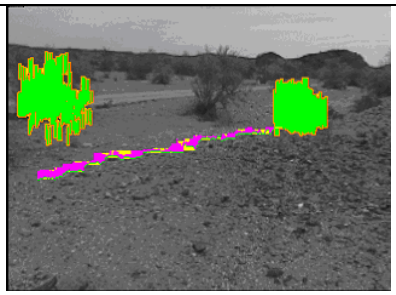
(d) Intensity image from Ft. Polk, LA



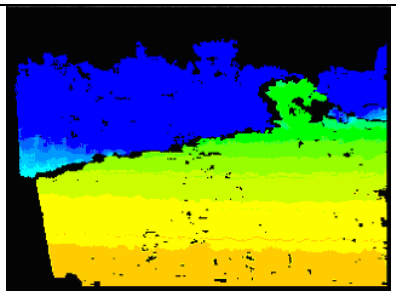
(e) Matching contours extracted from the scene in (d)



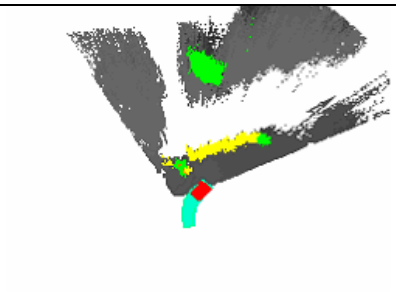
(f) Color coded tree hazard overlaid on the scene in (d)



(g) Intensity image of a drop-off at YPG with geometry based obstacle detection



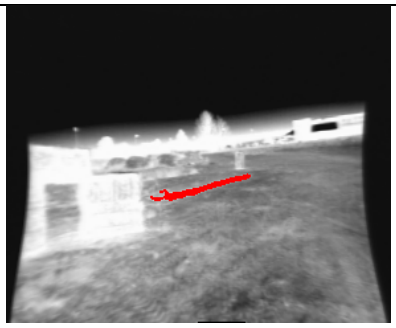
(h) 320x240 wide-baseline range image of the scene in (g)



(i) Raptor autonomously navigates away from the detected drop-off in scene (g)

(j) MWIR intensity image of a trench near SAIC, Littleton, CO

(k) Segmented, normalized intensity difference image for the scene in (j)



(l) Thermal signature based negative obstacle detection for the scene in (j)

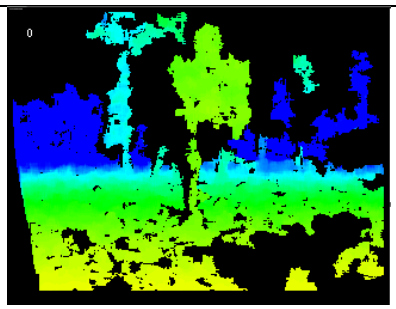
Figure 10: Sample Phase 3 passive perception results.

Three obstacle detection sensitivity modes were used during Phase 3: liberal, normal, and conservative. The OCU operator could select a sensitivity mode based on terrain characteristics. The liberal mode obstacle detection parameters were tuned for cluttered terrain, where there were no or few negative obstacles. This is the mode that would be used to navigate through a field of small bushes that are traversable. The normal mode obstacle detection parameters were tuned for fairly flat terrain that may contain trails and roads, few small bushes, and mild slopes and/or negative obstacles. This is the mode that would be used to navigate down a trial. The conservative mode obstacle detection parameters were tuned for terrain containing severe negative obstacles and slopes. The OCU operator also selected the daytime stereo baseline based on terrain characteristics. Narrow-baseline stereo was selected in cluttered terrain. Wide-baseline stereo was selected in open terrain. Mid-baseline stereo was rarely used. In practice, manual switching of baselines and sensitivity modes was tedious for the OCU operator. Here, auto switching software could be beneficial.

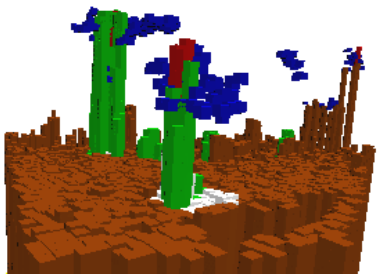
Figures 10d-f illustrates tree traversability analysis[4-6]. As shown in Figure 10e, an edge detection algorithm extracts long and vertical anti-parallel edge contours that correspond to the boundaries of individual trees. The stereo range data within tree trunk fragments are averaged and the diameter of each tree is estimated, based on the average range to the tree, the focal length of the camera, and the distance in pixels between matched contour lines. The traversability of each detected tree depends on its estimated diameter. In Figure 10f, the trees with red labels are a severe hazard and the tree with the yellow label is a moderate hazard. The tree hazard algorithm performed reasonably well during the daytime and nighttime. The contours algorithm, however, required the user to specify (in a parameter file) whether the trees were darker or lighter than the background. This required user input has subsequently been eliminated by running the contours algorithm twice, once with each setting, and concatenating the results. If there are range spikes or mixed pixels within a tree trunk fragment, averaging the range data within the fragment can result in skewed range and diameter estimates for a tree. From one cycle to the next, the location of a particular non-traversable tree can thus shift in the output terrain map. Further work needs to be done to improve the localization of non-traversable trees.

Figure 10g-i illustrates geometry based positive and negative obstacle detection. During autonomous navigation (with only passive perception running), Raptor detected the drop-off and navigated away from it. Figures 10j-l illustrates thermal signature based negative obstacle detection[7]. As shown in Figure 10k, MWIR imagery is segmented into closed contours, where each closed contour is assigned the intensity value that is the difference between the average interior intensity and the average boundary intensity. The intensity difference image is thresholded and filtering is applied (using stereo range data) to eliminate regions that geometrically cannot be a negative obstacle. The thermal signature based negative obstacle detector received limited field testing during Phase 3. This work could benefit from further modeling of trench temperature as a function of time of day, environmental conditions, and trench characteristics. In addition, further work is needed to constrain detection to structure within the perceived ground plane.

(a) Intensity image from Buffalo Creek, CO



(b) Color coded range image for the scene in (a)



(c) A 3D representation of a 30 meter terrain map for the scene in (a).

Figure 11: Sample Phase 3 passive perception results.

Figure 11 illustrates the Phase 3 representation for passive perception terrain maps. Passive perception terrain maps are north oriented, extend 30 meters, and contain 20cm x 20cm cells. Each cell contains elevation, terrain classification, obstacle, and confidence data. In Figure 10c, brown indicates the maximum ground cover elevation, blue indicates the minimum canopy elevation, green indicates the presence of a positive obstacle, red indicates the presence of a non-

traversable tree, and white indicates there is a low overhang obstacle above this cell. While multiple obstacle detectors can detect a single obstacle, only one obstacle type is reported for each cell. Only the maximum ground cover elevation is reported to TMAP. The cells containing low-overhang obstacles indicate where the canopy layer is too low for safe navigation. The current implementation does not do an adequate job of distinguishing low branches or the lower portion of a leaning tree from ground cover. Thus, the size of tree trunks is generally oversized.

6. CONCLUSIONS

The passive perception forward-looking sensor head and computing hardware performed reasonably well under a variety of environmental conditions (light rain, snow, summer and winter temperatures, dust, and smoke) at vehicle speeds up to 3 m/s. Although the sensor head total mass was slightly over the PTU payload limit, gaze control was fairly smooth when performed about the PTU home position. However, gaze control about the current steer angle caused the PTU to frequently break away from motor control, requiring a PTU reset (to re-home the PTU). In cluttered terrain, branches hitting the sensor head also caused the PTU to frequently break away from motor control. It is likely that a vehicle roll over and multiple PTU resets resulted in a poorly calibrated PTU, causing data blurring observed in the world map during gaze control on the last day of the YPG evaluation. Future designs should ensure sensors maintain calibration under harsh conditions. At YPG, the passive perception quad processor board failed in the field due to overheating and was replaced by a slower quad processor board. (During this time, the temperatures in the MBCU exceeded 115°F.) Some other mechanical issues that need to be addressed in future designs include keeping moisture (from wet vegetation and rain) off the optical windows and keeping the sun out each sensor's field of view, particularly during gaze control.

The passive perception system processed imagery at a resolution of 320x240 pixels, a fairly low resolution considering the mega pixel cameras that are currently available. Under the quad processor software architecture, the passive perception system outputted terrain maps at 2Hz/coprocessor during the A.P. Hill evaluation, and 1Hz/coprocessor during the YPG evaluation. (The drop in rate was due to additional software to perform terrain slope analysis, run-time data logging, and DIOC image serving at YPG). Low-resolution wide-baseline stereo ranging produced fairly dense range images on off-road terrain, except where there was a moderate to high amount of tall, thin vegetation. Thus, low-resolution wide-baseline stereo ranging performed poorly in detecting fallen trees and stumps partially occluded by tall vegetation.

The use of low-resolution narrow-baseline stereo ranging showed mixed results. In cluttered terrain containing moderate amounts of tall, thin vegetation, there was increased stereo matching that led to terrain maps containing more elevation data. And at times, serious hazards were only detected using low-resolution narrow-baseline range data (due to sparse wide-baseline range data in cluttered terrain). But because the range accuracy of low-resolution narrow-baseline range data is poor, obstacles in the narrow-baseline terrain maps tended to be severely oversized. And in scenes containing dense, tall, thin vegetation, stereo matching was poor. In practice, low-resolution narrow-baseline range data was useful only out to approximately 12 meters. At this range and image resolution, the range accuracy is approximately 1 meter.

To further improve stereo matching on tall, thin vegetation, narrow-baseline stereo ranging can be performed on high-resolution imagery (effectively reducing the angle each pixel spans). By reducing the pixel angular resolution from 3 mrad to 0.5 mrad, at a 2 meter range, the width of a 7x7 correlation window would be reduced from 4cm to less than 1cm. Processing high-resolution narrow-baseline imagery would also address the poor range resolution inherent to low-resolution narrow-baseline stereo ranging. The computation cost of stereo ranging would, however, significantly increase. Recent timing benchmarks on a 1.4GHz Pentium M laptop computer indicate that our correlation algorithm is the major timing bottleneck in stereo ranging. In processing 1024x768 imagery, the correlation algorithm accounts for as much as 85% of the stereo ranging cycle time. JPL is currently working on a field programmable gate array (FPGA) implementation of stereo ranging under a NASA funded project called Mobility Avionics Module. A drastic speedup in the correlation algorithm is anticipated. By solely adding a FPGA correlation engine, we project that stereo ranging could be performed on 1024x768 imagery at 6.25Hz (using the benchmark platform). This projection assumes a correlation processing time of 1 pixel per clock tick. Building FPGA implementations of the other stereo ranging components and using faster processors will further reduce the computation cost of processing high-resolution imagery.

One can also reduce computational cost by processing only a window of attention in the high-resolution imagery. In FY05, JPL will be implementing multi-resolution processing of high-resolution imagery under the U.S. Army Collaborative Technology Alliances program. Part of this work will involve fast processing of windows of attention on distant objects. If narrow-baseline stereo pairs are used for near range passive perception, and wide-baseline stereo pairs are used for mid and far range passive perception, then a fixed window of attention could be applied to the lower portion of the high-resolution narrow-baseline imagery, further reducing the computation cost of stereo ranging.

ACKNOWLEDGEMENTS

The research described in this paper was carried out by the Jet Propulsion Laboratory, California Institute of Technology, and was sponsored by SAIC under the DARPA PerceptOR program, through an agreement with the National Aeronautics and Space Administration. Reference herein to any specific commercial product, process, or service by trademark, manufacturer, or otherwise, does not constitute or imply its endorsement by the United States Government or the Jet Propulsion Laboratory, California Institute of Technology.

REFERENCES

1. Science Applications International Corporation, “Team Raptor PerceptOR Phase III Final Report”. Submitted to the Defense Advanced Research Projects Agency Tactical Technology Office under contract #MDA972-01-9-0015, Report #SAIC-PCTR-18A1-RaptorConfigMod.doc, March 31, 2004.
2. Bellutta, P., Manduchi, R., Matthies, L., Owens, K., and Rankin, A., “Terrain Perception for Demo III”, *Proceedings of the IEEE Intelligent Vehicle’s Symposium*, Dearborn, MI, October, 2000, 326-331.
3. Matthies, L., Bellutta, P., and McHenry, M., “Detecting Water Hazards for Autonomous Off-Road Navigation”, *Proceedings of SPIE Conference 5083: Unmanned Ground Vehicle Technology V*, Orlando, FL, April 2003, 231-242.
4. Huertas, A., Matthies, L., and Rankin, A., “Stereo-Vision Based Tree Traversability Analysis for Autonomous Off-Road Navigation”, *Proceedings of the IEEE Workshop on Applications of Computer Vision*, Breckenridge, CO, January 2005.
5. Huertas, A., Ansar, A., Matthies, L., and Goldberg, S., "Enhancement of Stereo at Range Discontinuities", *SPIE Defense and Security Symposium: Unmanned Ground Vehicle Technology VI*, Orlando, Florida, April 12-16, 2004.
6. Ansar, A., Castano, C., Matthies, L., "Enhanced Real-time Stereo Using Bilateral Filtering", *Proceedings of the 2nd International Symposium on 3D Data Processing, Visualization & Transmission (3DPVT)*, Thessaloniki, Greece, September 2004.
7. Matthies, L. and Rankin, A., "Negative obstacle detection by thermal signature", *Proceedings of the IEEE Conference on Intelligent Robots and Systems (IROS)*, Las Vegas, October 2003.