ARMY RESEARCH LABORATORY



RSTA for Small Rovers in Urban Warfare

Stuart Young, Pete Budulas, Philip Emmerman, Mike Scanlon, Nassy Srour, Dave Hillis, Phil David, Pete Fisher, Steve Vinci, Art Harrison, Kris Gurton, Sam Crow, and Mark Wellman

ARL-TR-1678 May 1999

Approved for public release; distribution unlimited.

AUTHENTICATED
U.S. GOVERNMENT
INFORMATION
GPO

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.

Army Research Laboratory

Adelphi, MD 20783-1197

ARL-TR-1678 **May 1999**

RSTA for Small Rovers in Urban Warfare

Stuart Young, Pete Budulas, Philip Emmerman

Information Science and Technology Directorate

Mike Scanlon, Nassy Srour

Sensors and Electron Devices Directorate

Dave Hillis, Phil David, Pete Fisher, Steve Vinci

Information Science and Technology Directorate

Art Harrison, Kris Gurton, Sam Crow, Mark Wellman

Sensors and Electron Devices Directorate

Sponsored by

DARPA

3701 N Fairfax Dr
Arlington VA 22203-1714

Approved for public release; distribution unlimited.

Abstract

The U.S. Army Research Laboratory (ARL) is pursuing a major research initiative in robotics. This research centers on collaborative physical agents that have advanced sensing, analysis, and behavioral characteristics that are linked to a mother ship that uses advanced visualization and awareness tools. Applications that are prompting this effort are reconnaissance, surveillance, and target acquisition (RSTA) of both human and vehicle targets as well as nuclear, biological, and chemical agent detection and localization. This report focuses on the requirements of a robot, or rover, to operate in urban terrain (such as military operations in urbanized terrain (MOUT) facility), to autonomously and stealthily approach enemy-controlled buildings, and to identify humans and any hazards to them. The requirements for this scenario could be performed by three increasingly complex systems, depending upon the extent of the operation. The three proposed systems are an individual agent, a team of collaborative agents, or a mother ship that works with a team of collaborative agents. This report focuses on a single physical agent solution. The agents must be able to negotiate all areas, such as curbs, stairs, and rubble, within an urban terrain. This report also discusses the application and component research thrusts of the RSTA module to detect humans and hazards to humans.

Contents

1. Introduction **1**

2. Perception for Reconnaissance Scenario **3**
2.1 Summary of Basic Behavior Modes for RSTA *6*
2.1.1 Move to Entry Point *6*
2.1.2 Cross Threshold *6*
2.1.3 Scanning Area (RSTA and Mapping) *6*
2.2 Summary of RSTA Processing in Scenario *6*
2.3 Summary of Events to Be Detected *6*

3. Point IR Detection and Cueing System **7**
3.1 Trade Space *7*
3.2 Design Solution: Pyroelectric Single-Element IR Sensors *7*
3.3 Background *9*

4. Detection of Suspended Trip Wires **11**
4.1 Trade Space *11*
4.1.1 Passive Detection *11*
4.1.2 Active Detection *11*
4.2 Solution: Use Laser Scanner and APS Camera *12*
4.3 Background *12*

5. Acoustic Detection From an Autonomous Vehicle **16**
5.1 Trade Space *16*
5.2 Speech Detection *25*

6. Infrared Camera for Mobile Urban Rover **27**
6.1 Trade Space *27*
6.2 System Requirements *29*
6.3 FLIR Selection *29*
6.4 Future FLIR Devices *30*

7. Image Processing for Reconnaissance **31**
7.1 Moving Object Detection *31*
7.2 Person Detection *33*

8. Future Directions **37**

9. Conclusion **38**

References **39**

Distribution **41**

Report Documentation Page **43**

Figures

1. Illustration of robot engaged in a mission 2
2. Illustration of robot in an operational urban scenario 3
3. Floor plan of a building to be searched corresponding to numbered steps in RSTA scenario 4
4. Pursuit deterrent munition-trainer 8
5. Line illumination raster scan pattern 14
6. Bow-tie illumination raster scan pattern 14
7. Wiper illumination raster scan pattern 14
8. Perpendicular translation illumination raster scan pattern 15
9. A bidirectional microphone schematic and photograph of Knowles and B&K microphones 18
10. Amplitude differences 19
11. Phase differences 19
12. Five-element directional array 19
13. Polar and Cartesian directivity curves from density plots 20
14. Polar and Cartesian directivity curves from density plots 21
15. Time series: talking, footfalls, and engine 22
16. Spectogram: time vs frequency and amplitude 22
17. Time-series data and spectograms for vehicle passing open offices with noise and speech 22
18. High-SNR speech: time (s) vs frequency (Hz) and amplitude (dB) 24
19. Time-series data for speech 24
20. Time-series data and spectograms of vehicle noise and data, unfiltered and band-rejected 25
21. Unfiltered, band-rejected data for sum of two harmonic representations 25
22. CCD and FLIR images of humans at 19 ft (CCD) and 14 ft (FLIR) 28
23. Detection of a moving person and his reflection in an IR image 33
24. Scatter diagram of blob perimeter contrast ratio vs number of lines required to approximate the blob's perimeter 36
25. Discrimination of people from other "hot spots" in IR imagery 36

Tables

1. RSTA processing in scenario 6
2. IR camera specifications 30

1. Introduction

This report focuses on requirements of a robot to operate in urban terrain. As discussed in the abstract, three proposed systems are being investigated for the scenario, depending on the extent of the operation. The first minimal system, which is the current focus, uses a single physical agent, or rover, that can be inserted into an urban setting to detect any human activity or hazards and report them to an operator. The second system consists of several (currently targeted to be four) similar physical agents that collaborate to clear individual buildings more quickly and efficiently or to clear different buildings simultaneously. An operator would also control this system with a handheld control. The agents can be sentries for one another, are communication relays, or share gathered information for multiple views of the same target. The final system consists of a manned mother ship, which deposits the four small robots near the area of interest. The mother ship will contain an advanced visualization operator control station with terrain, weather, local perceptions, and other team inputs. These four robots, with embedded biological, chemical, visible, infrared (IR), and acoustic sensors, will then perform an intelligent individual and collaborative search.

This small urban rover must reliably detect hazards (to humans) and do so with minimal unnecessary distraction (false alarms and unintelligible data) to the robot operator. The urban rover reconnaissance, surveillance, and target acquisition (RSTA) design must balance two conflicting constraints: having limited onboard sensing and processing resources and maintaining minimal false alarms. To achieve this goal, we selected a two-stage, multispectral sensing and processing approach. In the first stage, acoustic and point IR sensor arrays, which have low weight, power, volume, and requirements processing (with the acoustic array approach described in sect. 4), will act as cueing and coarse direction-finding (DF) devices. The acoustic array processing will additionally capture any sound above a set threshold, as well as detect voices, for the human operator.

The second sensing and processing stage uses IR and visible arrays and image processing. This more complex and expensive stage is necessary to keep the false alarm rate and subsequent communications and human workload at an acceptable level. The contrast increase from an IR array as compared to a visible array (either with natural light or strobed light) significantly increases the probability of detection and reduces the false alarm rate for automatic target detection. Another advantage of the IR array is the capability for long-range viewing at night. A strobed approach can only provide adequate light at short range. This is acceptable for navigation, but scenarios may exist in which being able to view at a distance greater than 30 ft is necessary. Figure 1 shows an illustration of a robot engaged in a mission.

The following sections discuss in detail the algorithms and processing for the RSTA. RSTA is also called perception for reconnaissance. The topics discussed are a proposed default scenario, the point IR detection system, acoustic detection, speech detection, IR imaging, image processing, and moving object detection.



Figure 1. Illustration of robot engaged in a mission.

2. Perception for Reconnaissance Scenario

A proposed default RSTA scenario (fig. 2) consists of the following steps with the corresponding RSTA activities. In this report, we present a default behavior. The operator has the ability to modifiy the sensing and processing depending on the mission. The following steps correspond to the locations in the floor plan shown in figure 3:

Step 1.

The rover traverses the outdoor terrain of the urban environment as it moves toward the target building. During the traverse, the following RSTA activities are occurring:

The acoustic array is continuously sensing to detect loud noises, gun-shots, voices, or any other noises that might indicate human presence. The array is not only sensing for noises but is also direction-finding to cue the visible or IR imager to the direction of the noise source. It will also transmit a 3-s audio clip to the operator when a voice is detected or when commanded by the operator. Additionally, during the traverse, the point IR sensor is used to sense whether or not an IR source has crossed the beam. Either a moving person crossing the stationary vehicle, the moving vehicle crossing a stationary person, or the moving vehicle crossing a moving person can trigger the point IR sensor. If any of these cause an alert, the visible imager or IR imager is cued to that direction and further processing is initiated to determine if the image contains a human. Also, the operator has the option of having a thumbnail image sent back to the control unit. The acoustic array and point IR sensors are always on to cue the other sensors.

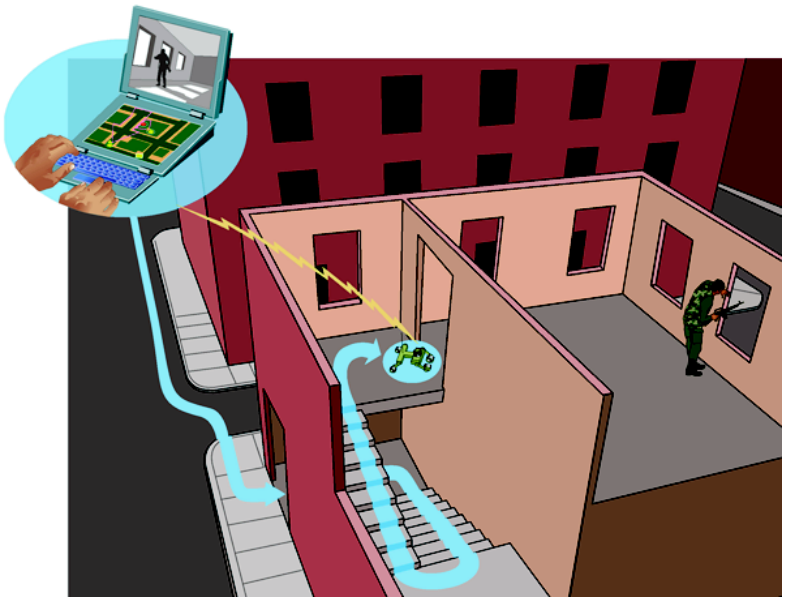


Figure 2. Illustration of robot in an operational urban scenario.

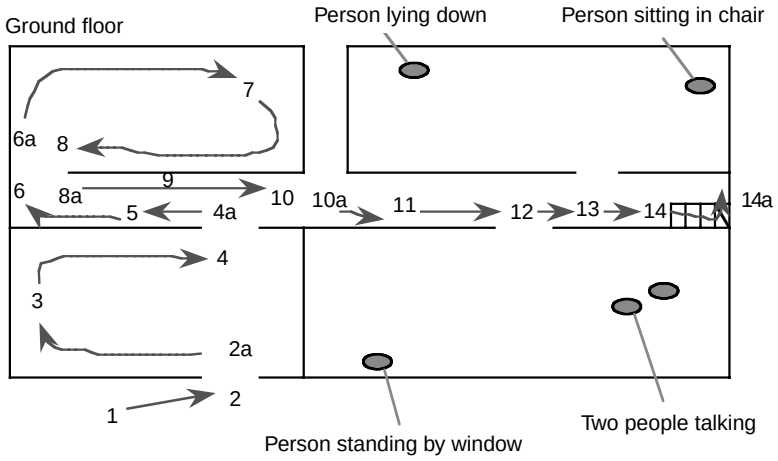


Figure 3. Floor plan of a building to be searched corresponding to numbered steps in RSTA scenario.

Because the IR imager takes so long to thermally stabilize, it is logically used while the vehicle is in motion for additional detection. During the traverse, the IR imager takes single-frame images and processes them by using a person-detection algorithm. The person-detection algorithm works while the robot is stationary or moving, and it views the image for hot spots that resemble humans. The IR imager can take one snapshot every few seconds or more often depending on the processor load. These images will normally be taken in the direction of motion, except when the IR imager is cued by other sensors.

Step 2.

When the rover approaches a threshold or opening, such as a doorway (determined by the navigation perception module), it stops before the opening and scans forward through the opening with either of the imaging sensors (from both sides of the opening and straight into the opening) and then turns around and scans in the opposite direction. The images are processed with motion-detection or person-detection algorithms. If a human is detected, the operator is alerted and a thumbnail image is transmitted to the operator. The operator then analyzes the image of the detected human to determine if it is friend or foe.

Also at the threshold, the rover will scan the opening for trip wires (in the future). The proposed method is to roll the body to the left and right while the scanning laser range finder is spinning. When the laser crosses the trip wire, the corresponding image is detected by the APS camera. Again, this is a topic of future research.

Step 2a.

The rover moves just inside the threshold. The rover scans the interior with the imaging sensors and then processes the images for motion, people, high-thermal contrast, windows, skylights, and doors. Upon entering, the rover sends the processed image to the operator. As before, the operator is alerted if any humans are detected.

Steps 3 and 7.

The rover moves about the room and scans for humans with the point IR, acoustic, and IR imaging sensors. Also during this time, it is constructing a map of the room with the laser range finder and stereo imagery.

Steps 4, 6, 8, and 10.

When the scan of the room is complete, the rover moves to another threshold and repeats the procedures as before.

Steps 4a, 6a, 8a, and 10a.

The rover crosses the threshold and repeats the actions in step 2a.

Steps 5, 9, 11, and 13.

The rover traverses the hallway and operates as in step 1.

Step 12.

The rover approaches a threshold where humans have been detected. Just as before, it scans forward through the opening (from both sides and straight into the opening) with imaging sensors. Then the acoustic sensor, the point IR sensor, or the IR imager cues the system. Once the cueing occurs, the front of the rover is oriented toward the triggering source (if not already pointed in that direction) and the images are processed with the motion-detection and person-detection algorithms. The default mode is for the rover to exit the room when humans are detected and return to its last known safe position to report to the operator and await further instructions. If no instructions are given, the default is for the rover to continue searching the building as before. The operator has the option of telling the rover to continue in rooms where humans are detected, move away from the threshold and report, move on to the next room, or exit the building expeditiously following the route in which it traveled.

Upon being cued, the rover is pointed toward the source and stops. If enough light is available, the wide-angle visible cameras on each side of the rover and the motion-detection algorithm are used to detect motion. If the light is inadequate, the vehicle is pointed toward the source and the IR imager is used along with the motion-detection algorithm. After the motion-detection algorithm processes the IR image, the person-detection algorithm is executed to verify the detection.

Steps 14 and 14a.

When the rover arrives at a flight of stairs (determined by the navigation perception module), the rover looks up with imaging sensors and then processes the image as in modes 2a and 4a. On each landing, it looks forward and backward as in modes 2 and 4. The rover scans continuously for trip wires. (For soldiers, stairwells are the most deadly areas within a building. Therefore, careful clearing of the stairwells and the top of the stairs is an essential task.)

2.1 Summary of Basic Behavior Modes for RSTA

2.1.1 Move to Entry Point

The rover is continuously sensing (acoustic direction-finding, point IR, and IR imager), cuing the imaging sensor and processor and sending the image to the operator upon detection of voices, gunfire, loud noises, and the corresponding direction of the source.

2.1.2 Cross Threshold

The rover is continuously sensing, detecting trip wires and scanning the view with the imaging sensor into and away from the threshold.

2.1.3 Scanning Area (RSTA and Mapping)

The rover is continuously sensing, scanning the area when directed by mapping software and scanning the area upon operator request.

2.2 Summary of RSTA Processing in Scenario

Table 1 reflects the sensor use for all the RSTA functions.

2.3 Summary of Events to Be Detected

The algorithms and sensors have been designed to detect humans who are standing in a room, by a window, or at the top of a stairway; walking down a hallway or in a room; talking in a room (two people); sitting in a chair; and lying down. Other events that will be detected include gunshots, loud noises, windows, trip wires, and vehicles.

Table 1. RSTA processing in scenario.

Sensor	Algorithm	Vehicle status	Method
IR imager	Person-detection	Stationary or moving	Uses a single frame to detect hot spots that resemble humans
IR imager or visible camera	Motion-detection	Stationary	Uses a few seconds of images to detect motion
Acoustic array	Simple-detection, used for cueing	Stationary or moving	Uses directional microphone arrays to determine direction and processes the data in the frequency domain to detect human speech
Point IR used for cueing	Simple-detection	Stationary or moving	Uses directional point IR to determine if hot source crosses the beam

3. Point IR Detection and Cueing System

The cueing system must search for human presence in its vicinity and produce a signal and a pointing direction for the more advanced and complex image processing. This signal will allow the robot to pause, train the IR imager in the proper direction, and switch into its "stationary detection" mode. (In the stationary detection mode, the robot is not moving, so the problems of platform motion and self-noise are minimized. The image and acoustic processing systems can use more efficient and accurate algorithms and temporarily grab the central processing unit (CPU) share momentarily relinquished by the navigation and mapping tasks.) If detection is confirmed, the operator is alerted; otherwise, the current task resumes after a 3-s search.

A cueing system is needed that can operate while the robot is moving or stationary, indoors or out; that can provide 360° coverage; and that can provide a minimal processing burden to the CPU. It must have a high probability of detection and an acceptable false alarm rate. Since the operator is not necessarily aware of the alarms from the cueing system, the primary consideration is to not slow down the execution of the mission.

3.1 Trade Space

The acoustic system provides some cueing but will not detect people who remain quiet. Ultrasonic sensors are limited in range and have degraded performance outdoors.

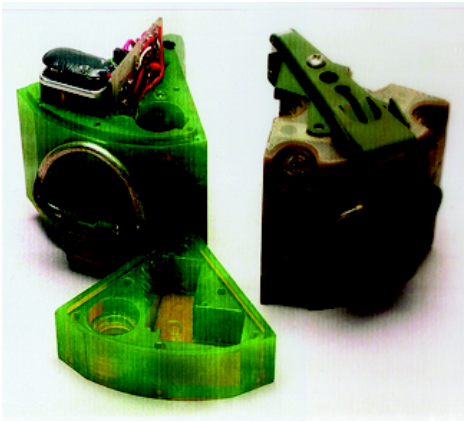
We evaluated a microwave motion sensor: MICROGUARD CS-95—a car alarm. During testing, it showed some attractive features, including a limited capability to detect motion around corners and behind objects. We ultimately dropped it, because no apparent way existed to adapt the sensor to operate from a moving platform.

3.2 Design Solution: Pyroelectric Single-Element IR Sensors

Our proposed solution is to create a sensor by using an array of four pyroelectric detectors. The design of the sensor is based heavily on the pursuit deterrent munition-trainer (PDM-T), a training device developed by the U.S. Army Research Laboratory (ARL).

The PDM-T (fig. 4) is designed to simulate a smart munition (mine). It sits on the ground and uses an array of four pyroelectric detectors to automatically trigger when a person walks past it. Although the PDM-T involves special-purpose hardware and packaging, some of the hardware and most of the algorithms are directly applicable to the robotic project. The PDM-T is described in some detail in the background section that follows.

Figure 4. Pursuit deterrent munition-trainer.



The PDM-T's detection algorithm can be applied with little or no change to the stationary or moving robot case. The rover will be moving at walking speed (about 1 m/s). The signature produced is much the same whether the person moves past a pyroelectric detector or the detector moves past the person.

The PDM-T was designed to operate in an outdoor environment while remaining still. Further, it was designed to mimic the probabilities of detection and false alarm of an actual mine. We conducted some informal, qualitative tests using a radio-controlled (RC) car to explore how placing the sensor on a moving indoor platform would affect its behavior.

We used two specially built test fixtures, closely modeled on the PDM-T. These fixtures are self-contained and allow testing of different parameter settings. For this system, the pyroelectric detectors would be sampled by an I/O board and the software would run on the rover's CPU. Each detector is sampled at only 10 Hz, so the processing burden involved is minimal. The four detectors were placed on the RC car in the configuration intended for the rover: one in front, one in back, and one on each side. The front and back detectors were intended to detect people crossing the path of the vehicle while the side detectors acted as hot spot detectors, searching for moving or stationary people whom the rover passed.

We tested the fixtures by driving the RC car through laboratory bays, offices, and hallways at the ARL facility. We followed with a handheld imaging IR sensor to visually examine the areas tested. The evaluation conducted was entirely subjective.

The false alarm rate was surprisingly low. When false alarms did occur, the IR imager usually confirmed that a real hot spot was present. These hot spots were caused by coffee pots, lamps, heating vents, and similar objects. Discrimination between humans and coffee pots is a higher level function that must be accomplished using the rover's IR imager or acoustic sensor.

The detection performance was uneven. Depending on the clothes a person wore as well as many other factors, some people had bright (hot) signatures and others had more muted signatures. People who had bright IR signatures were reliably detected up to about 15 ft or so, while their cooler (more fully clothed) peers were not.

The detection performance can be improved by using more efficient lenses and possibly by tilting the detectors upward so that they can scan the peoples' faces and hands. A simpler approach is to lower the thresholds in the detection algorithm and trade off a higher false alarm rate (which we can afford in a cueing system) for a higher detection rate.

3.3 Background

The PDM-T is a training device, developed at ARL in the early 1990s, that simulates the Army standard pursuit deterrent munition (PDM). This effort demonstrated an alternate target sensing technology (single-point IR) that could be applicable to future versions of the actual PDM, other trip-line function weapons, or scenarios requiring simple object detection. The detection and fire capability of the actual PDM are provided by trip lines that are ejected after an arming delay. The PDM-T uses four single-point IR sensors to simulate the trip-line function. These sensors allow easy reuse of the device, since trip-line triggering in a trainer would make each unit a single-use device. The IR sensors used are the Heiman Lhi954, which are sensitive in the 6- to 14-μm wavelength region and are ideal for human detection. Fresnel lenses were used to focus the IR in this region onto the detector surface. The PDM-T output signals from the IR sensors are paired, amplified, and filtered, resulting in two channels of low-frequency analog data. These signals are digitized and then processed by a target recognition algorithm. Upon detection of a valid target, the PDM-T produces both visual and audible cues to alert the soldier. The audio output is used to communicate with the soldier wearing MILES (Multiple Integrated Laser Engagement System) gear, which registers the kills produced by various training simulators.

The target recognition algorithm determines if a valid target is within the specified range and rejects most nontarget IR sources. By continually adjusting detection thresholds for ambient conditions and comparing input signal characteristics with those that are typically expected from known targets, the point IR array can achieve a good probability of detection and a low false alarm rate. The process begins when the device is activated, which begins a 1-min delay before the arming period. During this delay, the analog section is stabilized and the ambient IR conditions are registered and stored in memory. These values are updated continuously as conditions change. Detection thresholds are set lower during low noise ambient conditions and higher for noisy environments, such as those found during windy or rapidly changing temperature conditions.

This is a trade-off resulting in a somewhat reduced range in exchange for better false alarm rejection under such noisy conditions, while retaining the full detection range during quiet periods.

After the device adjusts for the environmental conditions, target identification and detection are then accomplished by looking for a specific type of change to the sensed IR conditions to occur. The known target signal profiles were obtained from extensive data collection of signals captured when a human subject passed through the zone of detection of the IR sensor. The scenarios used for this data collection included various walking speeds, clothing density of the subject, and sensor orientation in both indoor and outdoor environments. From this data, we obtained truly representative signal characteristics. This information was used as the basis for the signal comparisons used during the target detection process. The sampling rate for the analog data is 8 Hz, which is adequate considering that the signals from the detection process are in the 1-Hz range. The conversion, processing, and output signaling are performed using an 8-bit microcontroller unit (MCU) from the Motorola M68HC05 family. The P9 version of this MCU was used, which contains onboard A/D conversion, 2112 bytes of ROM, and an ultralow power sleep mode.

Technical testing of the PDM-T was performed by Test and Evaluation Command's (TECOM's) Electronic Proving Ground at Ft Huachuca. The results are summarized in test report EPG-TR-14-96. The performance of the IR detection was found to be quite good, with most criteria met or exceeded. The probability of detection for ranges from 0 to 5 ft and 6 to 10 ft were tested to be 97.4 and 96.8 percent, respectively. This test was performed on each sensor separately, with the other three sensors covered. Another performance assessment showed the absolute range of the PDM-T was from 6 to 23 ft, depending primarily on device orientation. When the sensor was aimed toward the center body mass of the target, increased range was recorded. In only 3 of the 56 trials of this test did the range fall below the expected value of 10 ft. This occurred when the sensor was aimed straight along ground level rather than aimed up toward the target. The other tests performed included safety, environmental, reliability, and human factors engineering. Most of the criteria for these tests were met and were related specifically on how the PDM-T behaved as a training device.

4. Detection of Suspended Trip Wires

To help reduce the risk to both military and civilian personnel, ARL is developing a class of robotics (both autonomous and remotely controlled) designed for use in various hazardous environments. One of the tasks that researchers would like the rover to perform is to seek and identify trip wires set in critical pathways that soldiers may encounter during missions.

4.1 Trade Space

4.1.1 Passive Detection

The natural illumination method for detecting suspended trip wires (favored among most robotic scientists, since it is the simplest to implement) involves applying various pattern recognition algorithms (PRA) to transmitted video imagery from visible cameras mounted on the robot. These PRAs are designed to "key on" and identify any "fine-line" structures that are present in the video scene. Unfortunately, this approach must overcome two fundamental problems. First, images of naturally illuminated three-dimensional (3-D) scenes do not convey the type of information necessary for PRAs to accurately distinguish between common straight edges (e.g., a sharp edge of a tabletop) and suspended wires. As a result, the false alarm rate is often quite high for all but the most simple of scenes. Second, by their very nature, trip wires are designed to blend into their backgrounds and thus often do not exhibit the necessary contrast needed for PRAs to key on. While the human handling involved in placing a trip wire produces a warming effect detectable in the IR, this signature is short-lived. Experiments using 3- to 5- and 8- to 12-μm IR imagers show little prospect of obtaining a reliable means of discrimination.

4.1.2 Active Detection

Active illumination methods (sometimes termed 3-D laser imaging/ Doppler) use a pulsed laser to illuminate an extended target by optically scanning a two-dimensional (2-D) area. Coincident sensors are then used to record the position and time delay in the scattered signal. A pseudo-image is generated that gives rough dimensions and distance to the illuminated object. These systems are complex and expensive to deploy. Furthermore, this technique is only effective in identifying targets that have reasonably large extended areas and is inefficient in identifying objects that possess small geometric cross sections, such as what is encountered with hanging wires or cables. Both passive and active techniques as outlined above are deficient in their approach because they cue

on features that are not entirely unique to the target of interest, i.e., trip wires and/or cables.

4.2 Solution: Use Laser Scanner and APS Camera

ARL's wire detection technique can be adapted to use the same sensors and image processing hardware already part of other systems on the robot. The wire detection technique (described in the background section that follows) uses a stationary camera and a laser that is rapidly scanned around its horizontal axis and slowly scanned vertically. The rover's navigation system includes a laser scanner and an APS camera that is sensitive to the laser's emitted light. The laser scanner rotates at a high rate in the horizontal axis. The actuators on the legs can easily produce the required, slow, side-to-side, rolling motion in the second axis. One of the APS cameras, which is offset about 1 in. from the axis of the laser scanner, can image the laser spot. (The 1-in. separation produces a parallax at 1 ft of 4.76°, and a parallax at 2 ft of 2.38°, which are easily resolved by the camera.)

Since the camera and laser scanner are both fixed to the robot's body, the rolling motion of the robot rotates the scanner and the camera. To simplify the image processing, the rolling motion is broken up into discrete steps and two images are captured at each step: one with the scanner on and one with the scanner off. Subtracting the two images produces a resultant image, highlighting the reflection of the laser scanner. This reflection sometimes includes an extended line that the laser scanner reflects from a distant object(s) (typically a wall). Any reflection from a suspended wire would be limited to a point and be offset from any other reflections by the parallax effect. The reflected points from a wire, at each rotation angle, would fall on a straight line in the coordinate frame of the camera.

4.3 Background

The reliable detection of trip wires has been a problem that has plagued the military research community for some time. Military and law enforcement agencies currently do not have an effective way of detecting simple trip wires. This problem is exacerbated during combat situations in which meticulous inspection of one's pathway is often not possible. Civilian law enforcement groups, such as the Drug Enforcement Agency (DEA), have reported an increase in so-called "booby-trapped" incidents involving their agents. Often when an illegal crop is identified, DEA personnel are placed at great risk during the secure phase of an operation in which booby traps are searched out and disarmed.

A similar problem involves the detection and early warning of power lines and hanging cables during certain helicopter missions. A particularly troublesome situation encountered by military pilots involves urban

night missions in which the probability of a helicopter colliding with a power cable or wire is greatly increased.

As mentioned before, ARL has developed a technique to automatically detect suspended wires by rapidly scanning a laser beam across a volume of space while examining that space with a video camera. Figures 5 to 8 show a series of generic schematics that outline the primary components. The schematics do not represent the only possible configurations.

We found three scan patterns that are uniquely suited for the illumination of suspended wires. The patterns were named to reflect the geometry of the resulting illumination: "bow tie," "wiper," and "perpendicular translation."

Acquiring each pattern starts with a common element consisting of a fine line of bright illumination that is created by reflecting an intense point source (in our case, HeNe, 0.6-μm laser) off a rapidly oscillating mirror; see figure 5. The oscillatory rate ω must be sufficient to produce what "appears" to be a continuous line of intense illumination. We found rates over 60 Hz to be adequate for video capture.

The next step is to alter the position of the "line illumination" in a time-dependent manner so that one of three possible scan patterns is created. Figures 6, 7, and 8 outline the mirror movement necessary to produce either the bow-tie, wiper, or perpendicular translation patterns, respectively. In each case, a much slower oscillatory frequency, ω_{slow} (i.e., $\omega_{slow} < 1$ Hz), is imposed around one axis of the oscillating mirror. This slow precession produces an extremely bright line of illumination. When projected on a plane surface, this line appears to pivot around the center point (bow tie), to pivot around one of the end points (wiper), or to translate a skewed line of illumination (relative to the wire being detected) in an up-and-down fashion (perpendicular translation). In figure 8, θ denotes the roll angles. One scan may have an advantage over another, depending on the specific application, but all three are designed to produce the same effect.

When a suspended wire or cable is in the illumination field of any of the three scan patterns, a bright "point" spot will result because of the light scattering off the wire at the intersection of the illumination line and the wire of interest. At some time, Δt, later, the point of intersection will have transversed a distance, $d(\omega_{slow})$, along the wire; reached an end point; and returned in the opposite direction along the same wire (see fig. 5). This type of illumination results in a distinct and unique pattern that is best described as a linearly moving point source that retraces its path in a slow, repetitive, oscillatory manner. Attempts to mimic this pattern using extended edges and geometries (i.e., table tops, chairs, metallic trim, etc) have shown it to be unique to suspended wires and cables.

Figure 5. Line illumination raster scan pattern.

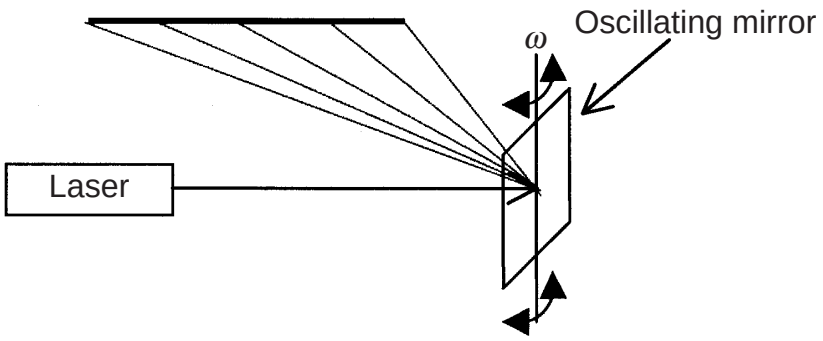


Figure 6. Bow-tie illumination raster scan pattern.

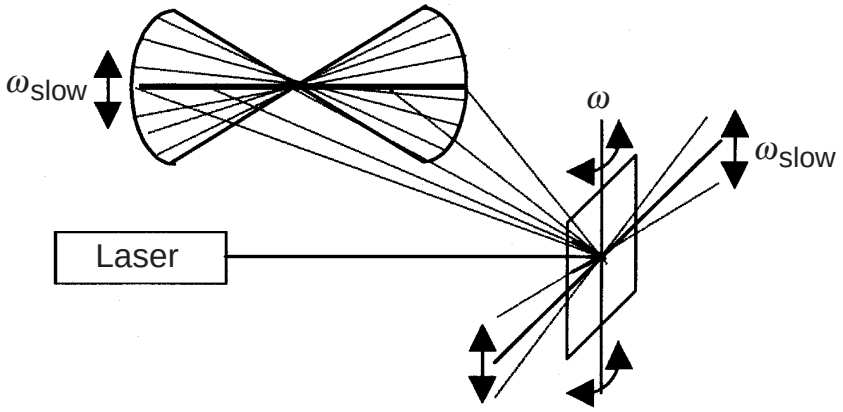


Figure 7. Wiper illumination raster scan pattern.

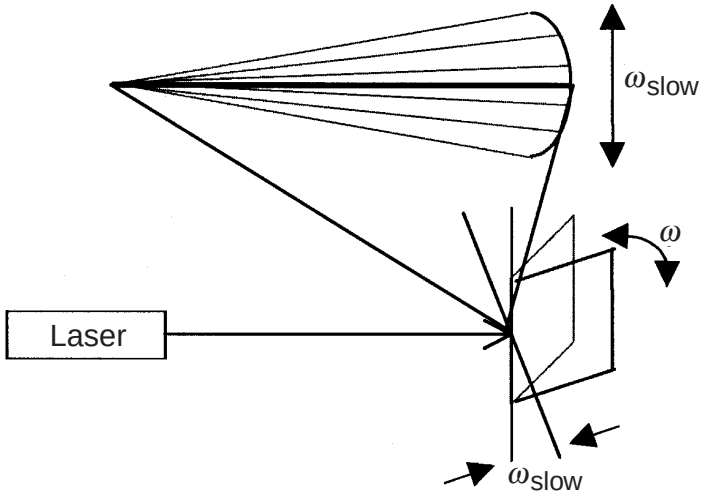
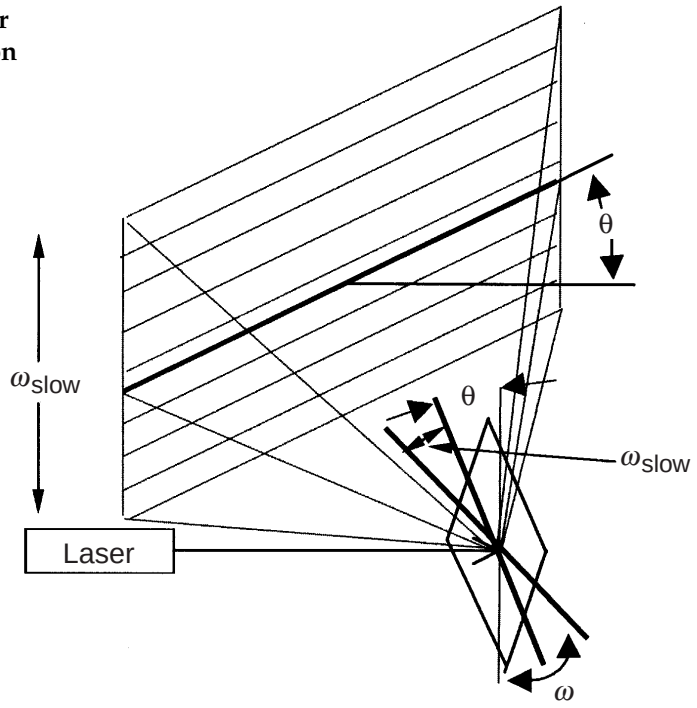


Figure 8. Perpendicular translation illumination raster scan pattern.



To best capture the resulting image for pattern recognition post-processing, a recording camera is slightly offset so that its field of view captures the linearly moving point reflection when a trip wire is encountered. Through the parallax effect, this point reflection is offset because of the reflection from a more distant extended object, such as a wall.

5. Acoustic Detection From an Autonomous Vehicle

The inclusion of acoustic sensors on an autonomous vehicle will add a new dimension to situational awareness and will augment and validate data collected from a vision detection system. Acoustics is a non-line-of-sight technology, which permits detections around corners, behind walls, or through obscurants. Signatures of interest include speech, walking, activity, weapon noises, number of people, TV, radios, and telephones. Automated detection algorithms of acoustic events can be further augmented by human auscultation.

5.1 Trade Space

Typical operational environments, which have reflective walls and floors, absorptive ceilings, hallways, doors, rooms, ventilation systems, and carpeting on walls or floors, can be acoustically challenging. Huge multipath reverberance can quickly become highly absorptive anechoic areas, confounding detections or localizations. Long corridors or hallways can channel sounds from great distances, which is good for detection, but bad for localization. Speech transmission through walls or doors can be beneficial, but may also limit or confuse.

Collecting acoustic data from moving platforms has always been a challenge:

- Motion induces vibrations of the microphone diaphragms.
- Structure-borne resonances and vibrations mechanically couple to the sensor through microphone mounting.
- Wind noise and turbulence can saturate high-gain amplifiers.
- Actual acoustic emanations from electromechanical mechanisms on the moving vehicle all contribute to the high dynamic noise.

This noisy condition is further compounded by hemispherical propagation of self-noise; radiated sound spreads in all directions, but the downward traveling sounds are again reflected upward from the nearly perfect reflective floor surface, effectively doubling the sound level from vehicle noise. To further complicate the signal-to-noise-ratio (SNR) problem, sounds occurring near microphones have a much greater effect than the same amplitude sounds emanating from farther away. Since atmospheric absorption in the spectrum of interest is negligible over relatively short distances, spherical spreading attenuates the signal 6 dB per doubling of distance. This essentially means that a small sound occurring near the microphones will be transduced as equal to a significantly louder sound occurring much farther away; hence, the “near-far” issue of self-noise and distant voice.

To improve the SNR of speech over vehicle, it is necessary to quiet the vehicle or create directional beams that are less sensitive to the vehicle noise. Quieting the vehicle can be accomplished by mechanical methods, such as acoustic insulation, damping, gear noise reduction, smoother tires, better shocks, or other methods that usually add weight and bulk to the vehicle. A more effective solution is to create highly focused zones of listening that are intentionally shaped to exclude the primary noise components of the vehicle. A more specific solution is four-directional microphone arrays. Directivity can be accomplished through the use of directional microphones or by the proper inphase combination of multiple microphones to form an array with maximum sensitivity in the phase-steered direction. A single omnidirectional microphone "hears" equally well in all directions. A single unidirectional microphone, such as a cardioid, has some sound reception preference in both the azimuth and elevation directions and rejects sounds approaching from the rear. An array of unidirectional microphones consists of several microphones usually arranged in a line of equal spacing, and the individual microphone outputs are summed simultaneously to produce a broadside directivity pattern with maximum sensitivity normal to the line. Only planar sound waves that approach the front of the array and hit all microphones simultaneously are constructively added inphase. Off-axis sounds, which do not traverse the array in a preferred perpendicular direction, are summed destructively producing phased cancellation and attenuation. Adding a second dimension to the line array, thereby creating a "planar" array, can further help reduce off-axis sounds propagating in the elevation direction as well as azimuth.

Having such a broadside array on the side of a vehicle will help eliminate sounds from the other three quadrant directions, as well as determine from which side of the hallway the speech is coming. The narrower the beam and the better the rearward sound suppression, the more likely the vehicle is to detect and locate speech in the direction normal to the array. Obviously, front, left, and right arrays can be compared to relate between signal strengths, frequency content differences, and time of arrivals to the location of targets. Combination of the arrays can create other beam patterns that might bisect the primary quadrants, such as left-front or right-front. Another option, not evaluated at this stage of development, is to use several omnidirectional microphones in a volumetric (3-D) array, and continuously steer or scan the beam in all or preferred directions by varying the phase delays between the microphones. (Sound speed, sensor separation, and steered angle determine the appropriate delays.) Noise cancellation techniques can be used by using a reference microphone near vehicle noise sources, which will become the reference standard by which noise and signal can be compared and separated.

Shown in figure 9 is the schematic and photograph of a Knowles EL-3077 bidirectional microphone with tube extensions. By the addition of

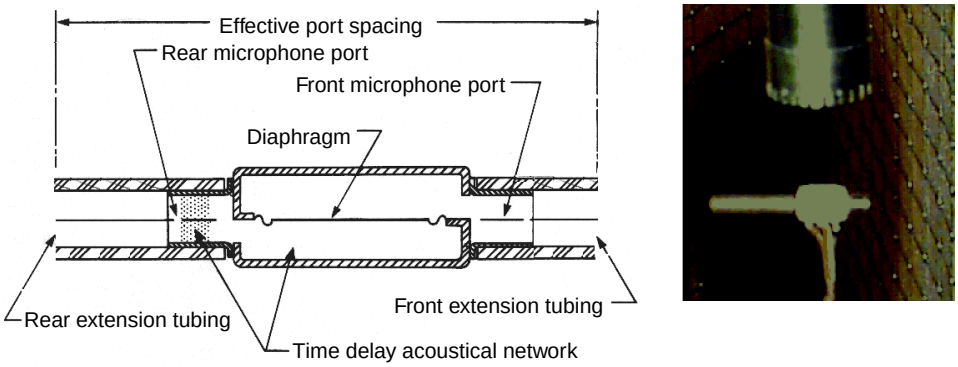


Figure 9. A bidirectional microphone schematic and photograph of Knowles and B&K microphones.

appropriate tube lengths to the front and back ports, a true cardioid pattern can be attained, of which its polar response is $1/2(1 + \cos \theta)$. Also in the photograph above the Knowles microphone is a Bruel and Kjaer (B&K) 1/2-in. 4147 microphone, an instrumentation grade omnidirectional microphone used as the reference for the array directivity measurements.

Because the effectiveness of the array relies upon the proper phase combination of multiple signals, it is imperative that the microphones chosen be properly matched to have similar phase and amplitude responses. B&K microphones have exceptionally flat amplitude and phase responses, but are prohibitively expensive and fragile. Using lesser quality microphones with nonlinear responses is common practice as long as each microphone responds in the same way. An assortment of Knowles microphones was tested to evaluate repeatability. The normalized frequency response is shown in figure 10, with the vertical axis representing decibel-volt differences between microphone outputs immersed in the same sound field. The normalized phase diagram in figure 11 has vertical units of degrees. The horizontal axis represents frequency in hertz. Twenty-one microphones are shown.

Ideally, but unlikely, all the microphones would have the same shape and the above curves would overlay themselves perfectly. To optimize similarity between available microphones, we chose three groups of the closest five microphone groupings for the three arrays, attempting to use both phase and amplitude response as grouping parameters.

The photograph in figure 12 shows a broadside cardioid array with each of the five elements 1 in. apart and the B&K reference suspended above the array fixture. We mounted the entire array fixture on a computer-controlled rotating table within an acoustic anechoic chamber. Data acquisition and table orientation were controlled with LabView software, and digitally sampled waveforms stored for postprocessing by LabView as well. By simultaneously exposing the array and reference microphone to high-amplitude broadband noise created by a distant speaker also in the chamber, we could fully characterize microphone response and directional response as the table was rotated.

Figure 10. Amplitude differences.

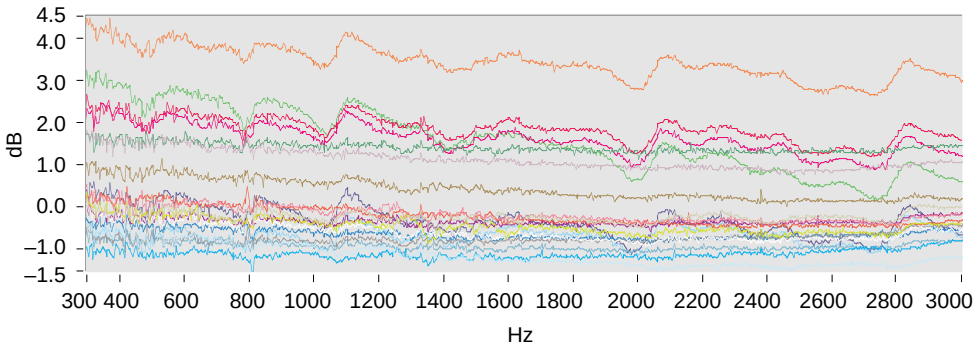


Figure 11. Phase differences.

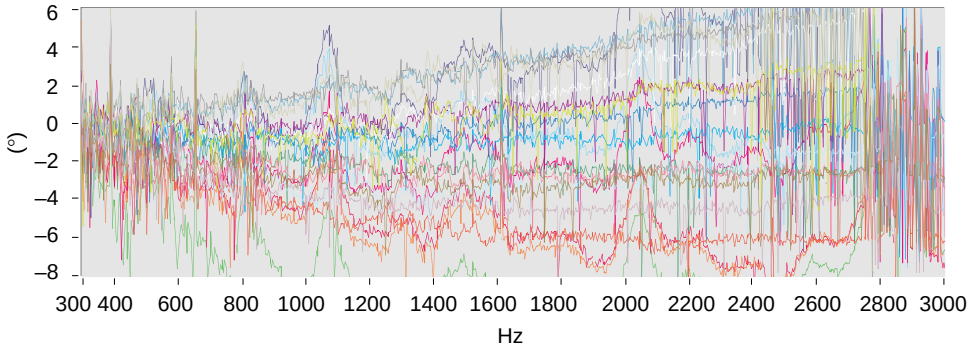


Figure 12. Five-element directional array.



By performing fast Fourier transforms (FFTs) on both the incrementally rotated array and the stationary reference microphone, we found that the resulting transfer function between the two sensors indicates the directional response of the array as it turns away from the sound source located at 0° azimuth. Figure 13 shows directivity representations of a single EL-3077 microphone without extensions, the same microphone with tube extensions to create a cardioid, and (as previously mentioned) a five-element array of cardioid microphones spaced 1 in. apart.

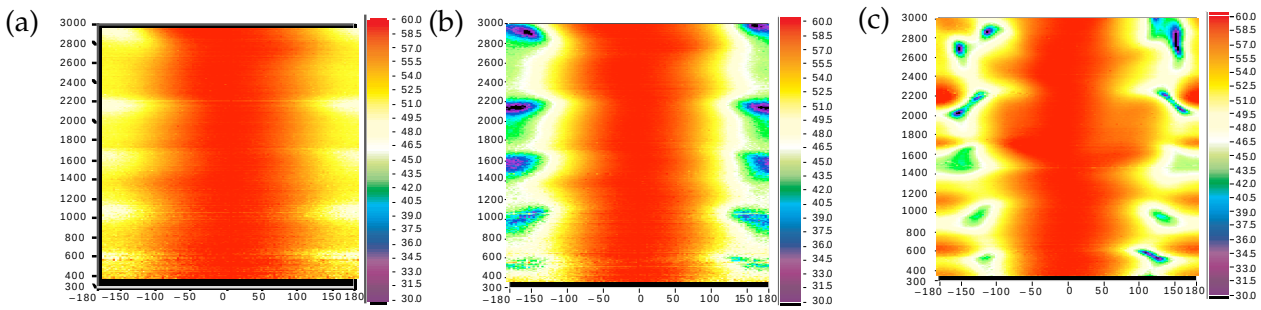


Figure 13. Polar and Cartesian directivity curves from density plots (a) microphone only, (b) single microphone with plastic tubes, and (c) array with plastic tubes.

The vertical axes in the graphics in figure 13 represent frequency of interest in hertz, and the horizontal axes are the array's rotation orientation in degrees azimuth, showing the center (0°) pointing at the speaker and normalized to an arbitrary sound pressure level of 60 dB. As the array is turned ±180° from the center, the resulting amounts of attenuation are determined by the color bar. Ideally, the present goal would be to create as narrow a beam as possible, with the greatest rearward attenuation.

To better understand the plots (fig. 13), one can take a 2-D "cut" at 2100 Hz, and graph it in polar coordinate fashion, where the microphone's forward direction (0°) is oriented toward the right side of the graph, and where the radial divisions represent 10-dB attenuation. Figure 14 (a) is not very directional at that particular frequency; whereas, figure 14 (b) indicates better rearward attenuation, but an extremely broad beam (gradual rounding). The plot in figure 14 (c) shows a slightly elevated rear lobe, but excellent rearward reduction and narrower frontal beamwidth (sharp roll-off). The larger rear lobe results from the microphone's separation being related to a fraction of that frequency's wavelength, creating suboptimal cancellation.

We should also like to mention that the microphones must be mounted so that the acoustic wave fronts traverse the array without being disturbed, from whatever direction they occur, so that the cardioids and the array can properly perform the direction-dependent enhancement or cancellation.

We conducted some preliminary experiments using data collected with a modified hobbyist's RC car. The RC car is extremely noisy when running and represents a signal-to-noise problem greatly over any acoustic level projected for the robot (based on acoustic measurements of some subcomponents). While much of these data represent something of a pathological case (one cannot really imagine this RC car in a stealthy mode), the data still provide insight into the problem.

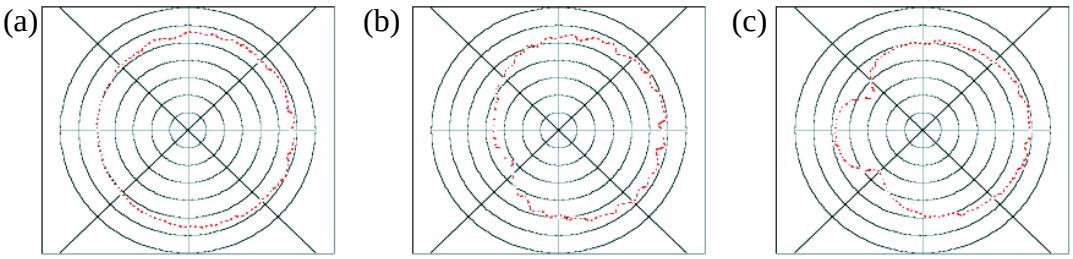


Figure 14. Polar and Cartesian directivity curves from density plots (a) microphone only, (b) single microphone with plastic tubes, and (c) array with plastic tubes.

We mounted three acoustic arrays on the three upper edges of an open-ended box and secured noise-deadening leaded foam atop the vehicle. The engine compartment also contained some foam; however, little was done to quiet the vehicle. The vehicle was tested in a typical laboratory hallway, which contained many laboratory bays and offices, two of which had people reading a script during the vehicle's run. Shown in figure 15 is a time-series plot and the resulting time-frequency spectrogram of data taken from the Sony digital audio tape (DAT) recorders located on the vehicle beneath the array. The data shown represent a "vigilance mode" in which the vehicle is listening for a target.

The forward array was pointing down a hallway at a range of 25 ft (1- to 2-s region) listening to a person talking. The person moved to 37 ft and talked to a person inside an office (2- to 4-s region), and then ran 12 paces toward the array (4- to 9-s region). Between the sound of the running person's footfalls, the spoken words of the people in the offices were distinguishable over ambient noise.

The spectrogram in figure 16 demonstrates that acoustic detection of individuals is easy to do while in the vigilance mode and at distances much greater than shown in this report. At the 10-s point, the vehicle was turned on and began moving. Note, however, the high-sound levels of the vehicle traveling at 69 ft/min, nearly 40 dB above the hallway ambient. From this spectrogram, it is clearly visible that the vehicle's self-noise and the speaker's voice contain similar frequency components and temporal (impulsive) structures. This similarity causes voice detection to be difficult while the vehicle is moving, requiring enhanced directivity to overcome the noise.

Figure 17 shows three array outputs as the vehicle traveled at 69 ft/min past an open door of an office in which a person was reading aloud at a normal speaking level. The person speaking was on the right side of the vehicle as it passed the doorway, and directly across the hallway, on the vehicle's left, was a large noisy computer apparatus just inside the laboratory bay, which created acoustic interference noise. The time-series data

Figure 15. Time series: talking, footfalls, and engine.

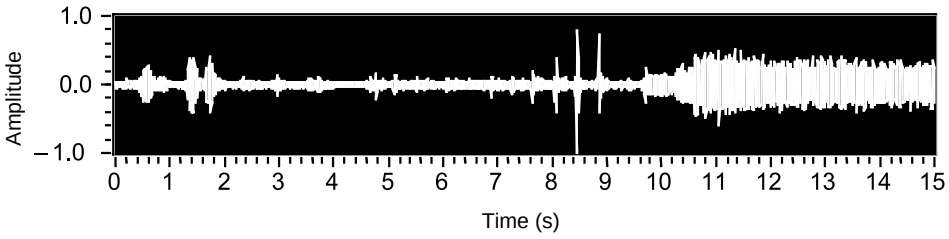
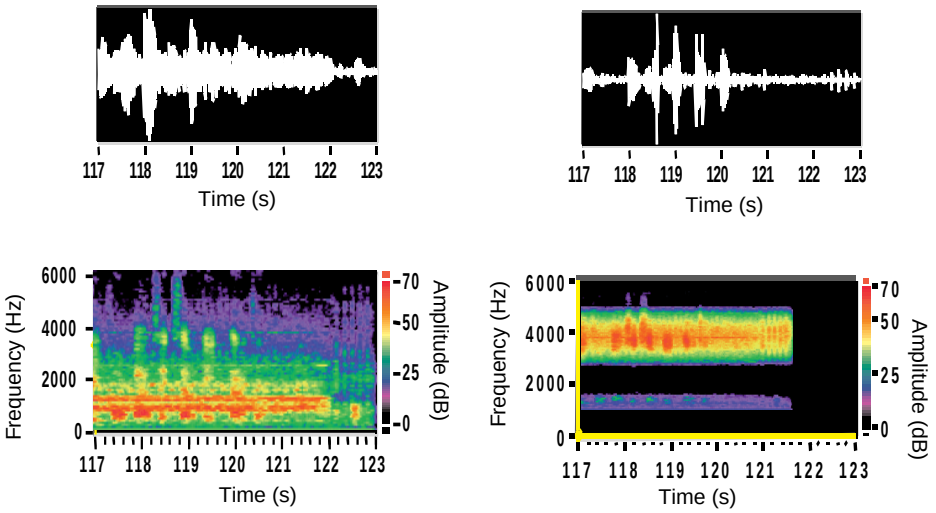


Figure 16. Spectogram: time vs frequency and amplitude.



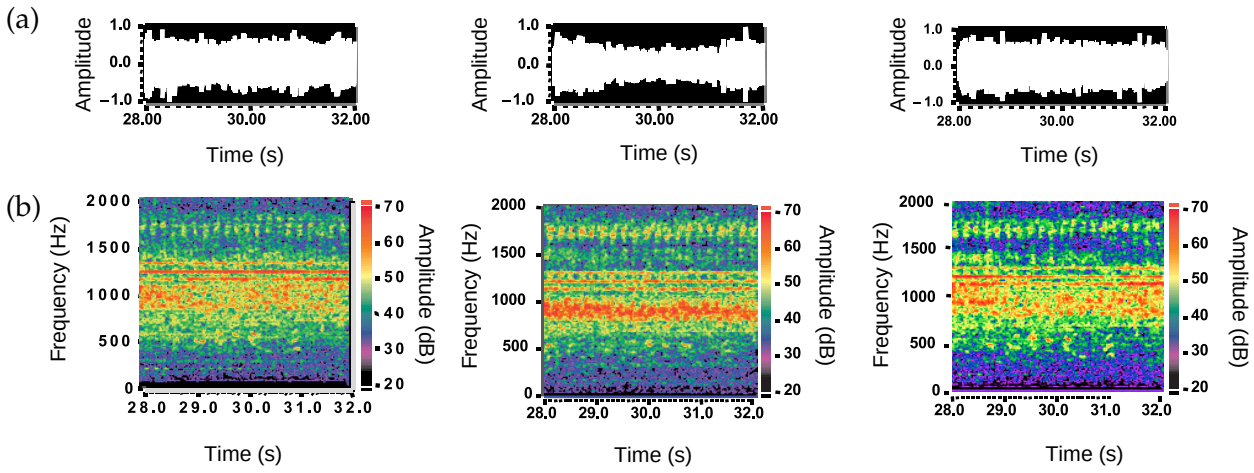


Figure 17. (a) Time-series data and (b) spectograms for vehicle passing open offices with noise and speech.

from the left array shows consistently high broadband noise, since the array was constantly pointing in the direction of the noise source throughout the 4-s run. The data from the right array were able to remove some of the interfering noise, as seen in the reduced midsection of the time data and in the lower amplitudes in the spectrogram. The speech components, as seen primarily below 800 Hz in this data set, are slightly higher in the SNR as well. The front array has consistent levels throughout the segment and hears some components of the speech and interfering noise.

Voice detection methods are numerous and complex. Sensor systems and detection algorithms have great difficulty exceeding the sensitivity of the human ear and the mind's ability to interpret subtle differences in frequency content, relative amplitudes, phase and time of arrival differences, as well as accurately filter noisy data to extract speech. Data recorded during vehicle motion were extremely cluttered with distracting noise but were discernible with the naked ear when played back through a speaker (or if listened to remotely during future operations). A combination of band-pass and band-rejection filtering significantly helped remove vehicle noise and enhanced the speech SNR so much that speech was clearly heard. However, an automated detection algorithm in a dynamically changing noise field is not easy. Of primary significance in voice detection is harmonic component analysis relating to the speech formants.

Shown in figure 18 is a zoomed-in version of the vigilance mode speech data (vehicle stationary) seen previously. The harmonically related components of the spoken words with high SNR are seen in red and orange. Note also that the harmonics are frequency-modulated depending on which sounds are uttered. Temporal and spectral cues are currently used to detect and interpret speech for voice recognition and voice commands.

As depicted in figure 19, the time-series data of the first word (between 0.5 to 0.6 s) in the spectrogram show repeated patterns. By visual inspection of the waveform in figure 19, approximately 12 bundles of speech can be seen in the 0.1-s window. In 1 s, or a 120-Hz fundamental, with higher harmonics also visible, approximately 120 bundles would be derived. Remember that a 300-Hz high-pass filter in the preamplifier circuit was intentionally chosen to suppress all lower-frequency vehicle sounds and nearby traffic, equipment, and ventilation noises. Unfortunately, this filter also attenuates the fundamental and first harmonic of this particular speaker's words. It may be beneficial to lower the high-pass filter corner to 90 Hz to include most people's lower fundamental and several higher harmonics for detection ease. Lower frequencies, by the way, tend to travel the farthest in atmosphere, and by nature of their larger wavelengths, tend to wrap around objects and go through structures better than higher frequency sounds do, making detections easier.

Much of the dominant vehicle noises were either high levels of broadband, tonal, or impulsive in nature, but were not necessarily uniquely harmonic in structure. Speech, on the other hand, always contains harmonically related and frequency-shifting components. As shown in figure 20, a harmonic analysis of unfiltered and filtered data was compared for the presence of harmonics by essentially calculating and summing all the harmonic values taken from each power spectrum as time progressed. For each fundamental, the harmonic summation used the center frequency from each of the adjacent FFT frequency bins between 3 and 300 Hz and calculated harmonics for the entire bandwidth.

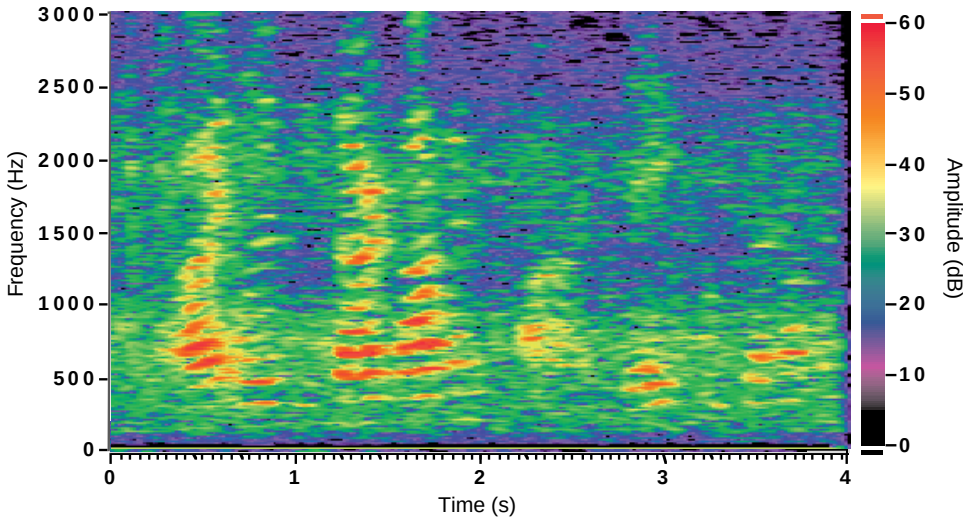


Figure 18. High-SNR speech: time (s) vs frequency (Hz) and amplitude (dB).

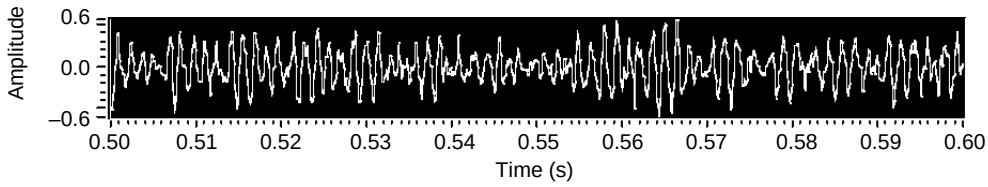


Figure 19. Time-series data for speech (0.1 s shown).

The top two graphs (fig. 20) represent the time-series representations of the unfiltered and filtered data, left and right, respectively. Note the enhanced SNR on the filtered data—it clearly shows the five words spoken between 118 and 121 s. The filter chosen is a 31-tap band-rejection filter that eliminates the 800- to 2200-Hz region. We calculated the two spectrogram representations (fig. 20) using 2048-point Hanning FFTs with 90 percent overlap on data sampled at 12 kHz. The two harmonic representations are shown in figure 21. Note how clearly the filtered harmonics indicate speech; whereas the unfiltered version on the left does not discriminate because of the broadband noise that can contain randomly high values, which contribute to certain harmonics. These data support the need for a quieter vehicle but also demonstrate that an automated harmonic analysis algorithm can detect speech while a vehicle is moving at top speed.

We have shown that voice is detectable over the vehicle's dynamic noise by band-pass filtering and harmonic analysis. It is also detectable by the human ear listening through headphones, which may become an option for an operational system. The directivity of the array was acceptable for this initial vehicle evaluation but must be enhanced to improve localization and detection and remove extraneous sounds and vehicle noise. An automatic gain control circuit should be implemented to lower the gain during transit and increase sensitivity when the vehicle is paused in the

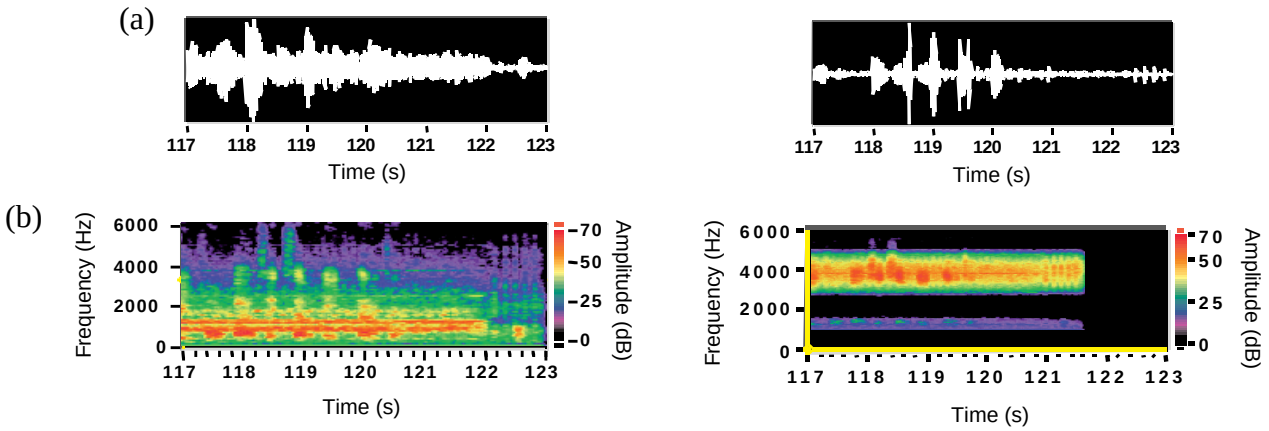
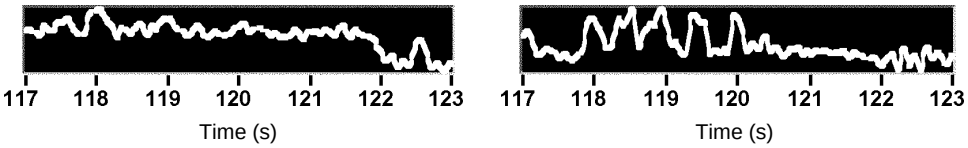


Figure 20. (a) Time-series data and (b) spectograms for vehicle noise and data, unfiltered and band-rejected.

Figure 21. Unfiltered, band-rejected data for sum of two harmonic representations.



vigilance mode. Better adaptive filtering and speech detection algorithms also must be implemented. Obviously, we could have used quieter vehicles for this experiment. However, we believe that this unimproved embodiment might be a "worst-case" experiment to challenge all aspects of design and signal processing. We were successful. The lessons learned and the experience gained from these accomplishments will only help to further acoustic advancements more easily, using improved hardware.

5.2 Speech Detection

Speech can be characterized as a nonstationary process with great variations in the short-time power. Drago et al (1978), modeled the speech signal as two nonstationary random processes: one was band-limited between 300 and 3400 Hz, and the second, primarily because of the formants in speech and slow time-varying envelope of speech, was between 0 and 50 Hz. Using this model, they were able to construct robust speech detection algorithms. However, work by several researchers, including William C. Newman (1990), shows that the main formant of speech is band-limited from 0 to 300 Hz and was the major feature that an adaptive neural network used for speech detection.

Any speech detection algorithm needs high-noise immunity and endpoint detection of the speech signal, and it must consider the particular

characteristics of noise to differentiate speech. Noise in the channel is primarily because of stationary noise (both vehicle and surrounding sporadic noise), nonstationary noise (nearby activity, doors opening and closing), and other sources. Sporadic noise is too complex to contend with and therefore should not be considered in the algorithm design because of unwarranted complexity (Taboada et al, 1994; Turk and Pentland, 1991).

We propose a simple speech detection algorithm. The data are band-limited from 0 to 300 Hz first by the use of a multirate filter architecture, which ensures low-complexity filters and improved resolution of the frequency bins. The 0- to 300-Hz region also provides a great deal of noise immunity and, as stated earlier, is the main formant region. The algorithm then will identify the short time-peak "power" for each 125-ms time block. This time slice is favorable (Drago, 1978) for simple speech detection, since shorter slices will cause more interruptions in detection because of the presence of unvoiced speech sounds or weaker speech signals. Short-time slices would be required if we intend to perform speech recognition (Waibel et al, 1989; Makino and Kido, 1996). Also, longer time slices will tend to average out the dynamic speech feature, which assists in determining speech onset. The ambient noise will be determined and a subsequent set of thresholds for use in a band-crossing algorithm. Speech will be detected if the short time-peak power exceeds a maximum threshold and the dynamic feature, which is the ratio of the short time-peak power that (in the previous 125-ms time slice) exhibits a peak. This dynamic feature exhibits a sharp peak at the onset of speech and, to a lesser extent, when the speech subsides. Once speech is detected, the band-limiting thresholds will be used to determine if speech is still present, and speech will be considered absent only when the total power falls below the lower threshold for two 125-ms time slices. This algorithm incorporates several positive features of algorithms mentioned in the literature of Drago (1978), Newman (1990), and Taboada (1994).

6. Infrared Camera for Mobile Urban Rover

One aspect of the mobile urban rover (MUR) mission is the detection of personnel inside and outside buildings of interest. Acoustic and point IR sensors will be used on the MUR as trigger sensors to indicate possible human presence. Data from these sensors will give a relative direction of the detected possible human presence. The robot then needs a sensor subsystem that is capable of scanning those given areas to confirm the human presence. Such a sensor subsystem must be capable of providing high-contrast imagery over a wide range of lighting and obscurant conditions in indoor and outdoor situations. This imagery would then be processed by the computer located on the MUR to verify the presence of humans.

6.1 Trade Space

Imaging sensors available for detection and tracking of personnel include CCD cameras, image intensifiers, and infrared cameras. The MUR imaging sensor has the following requirements:

- Absolute maximum weight of 12 oz, 6 oz desired.
- Low-power consumption (battery-operated).
- Small size, space limited to $3 \times 3.5 \times 2.5$ in.
- Provide images in indoor and outdoor environments.
- Provide images under no-light to bright-light scenarios.
- Depth of field from 5 to 50 ft.

While CCD cameras provide a small, lightweight, and low-power solution for this application, they are unable to provide proper operation in low-light and no-light situations. Their operation is further degraded if smoke or obscurants are present. Examples of images from CCD cameras can be seen in figure 22 (a) with the room lights on and figure 22 (b) with the room lights off. With the lights off (fig. 22 (b)), insufficient information is available to identify the subject in the room. Infrared illumination sources can be used to enhance the operation of the CCD camera when room lighting is insufficient. However, these sources have a limited range of operation. Figure 22 (c) shows an image taken with an IR illuminator and CCD camera with the IR filter removed. In this case, the illuminator is a light-emitting diode (LED) based device with a wavelength of 880 nm that consumes 8 W of input power. Figure 22 (d) is the same image with the contrast increased by 95 percent. This image shows that some information is obtained using the IR illuminator. Inspection of the IR illuminated images by ARL staff indicates that moving object detection and tracking are possible using this imagery, but that performance (i.e., probability of detection and probability of false detection) will be reduced

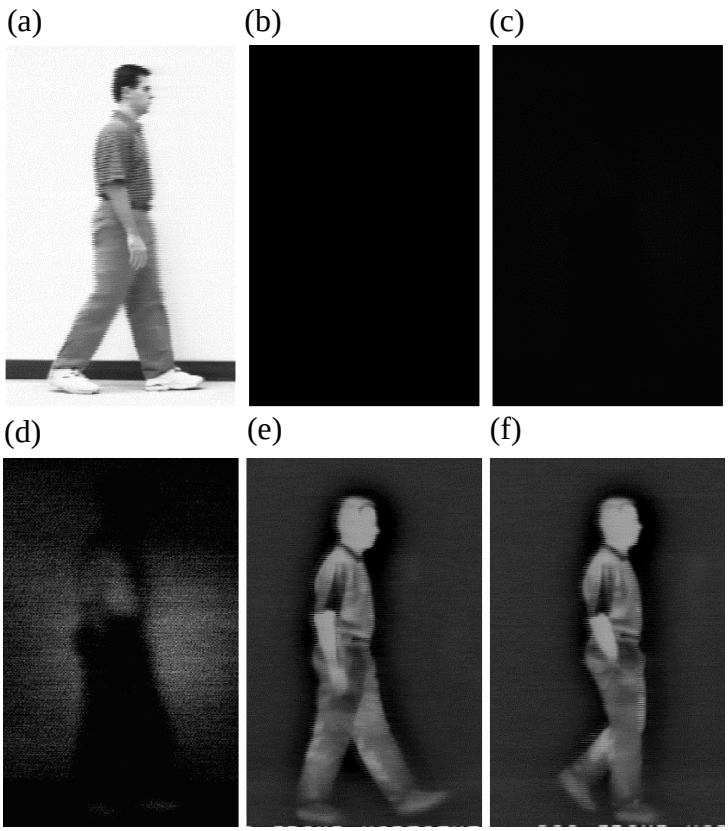


Figure 22. CCD and FLIR images of humans at 19 ft (CCD) and 14 ft (FLIR): (a) CCD with lights, (b) CCD without lights, (c) CCD with IR illuminator, IR filter filter removed; (d) CCD with IR illuminator, without filter, 95 percent increase in contrast via software; (e) FLIR with lights; and (f) FLIR without lights.

when compared to the same processing applied to IR imagery. The reduction in performance is due to low-image contrast as compared to the noise level in the image. An additional complication with using IR illuminators is the chance that personnel wearing night-vision equipment may detect the IR source.

An IR strobe could be used in place of the IR illuminator for low-light CCD camera operation. Problems associated with the application of an IR strobe include limited viewing range and limited frame rate.

Image intensifiers could be used to provide imaging capability under low-light conditions. These cameras require some minimal light to operate and cannot operate in no-light conditions. In addition, image intensifiers suffer image washout if bright sources, such as windows, skylights, or distant lights, are present in a darkened room.

IR cameras (FLIR) are imaging devices that produce an image in which the intensity is directly related to the surface temperature of the object being viewed. Such an image does not depend on a local light source as indicated in figure 22 (e) (room lights on) and 22 (f) (room lights off). These images provide sufficient contrast for detection and tracking of humans under a wide variety of situations. The range of operation for an IR camera is limited only by the detector resolution and the optics used.

In addition, these cameras can see through smoke and many other obscurants in situations that a CCD camera or image intensifier might be useless.

Note that most IR cameras have a relatively small instantaneous field of view (FOV), typically 40° to 50° for a near-depth-of-field lens as would be used in a MUR-type application. This precludes the possibility of using the IR camera as the only personnel detection sensor. The combination of wide FOV trigger sensors (acoustic and passive IR) with a limited FOV confirmation sensor (IR camera) provides a complete set of sensors for the detection and confirmation of humans.

6.2 System Requirements

The following specifications are given in the selection of a FLIR for the MUR vehicle:

- Maximum weight of 12 oz, 6 oz desired.
- Maximum package size $3 \times 3.5 \times 2.5$ in.
- FLIR operational with 1 min of power up.
- Less than 8 W.
- Operation over range of 5 to 50 ft.

The following derived specifications are set by mission needs:

- "Uncooled" technology to speed start-up (8- to 12-μm operation).
- Lowest power consumption possible (battery operation).

6.3 FLIR Selection

ARL staff investigated a large number of FLIR cameras for potential application to the MUR program. Of those systems currently available, only two almost meet the requirements of this project. All other systems far exceed the weight, size, and/or start-up times required by the MUR. The two candidate cameras are the IR Microcam and the Texas Instruments IR camera core used in the VideoTherm 2000. Specifications for these cameras are listed in table 2. The IR Microcam has been selected as the FLIR for the MUR as it is the only unit currently available that meets the minimum height requirements to fit the camera components into the MUR body. Future FLIR systems will provide substantial reduction in weight, size, cost, and power consumption as indicated in the next section.

6.4 Future FLIR Devices

Several companies are currently developing FLIR camera systems that will provide reduced weight, size, and power consumption at the cost of reduced detector array size. Raytheon TI Systems is currently developing a 160 × 120-pixel microbolometer-based sensor for a Defense Advanced Research Projects Agency (DARPA) program. In addition, Indigo Systems is developing a 160 × 120-pixel microbolometer-based sensor for a Communications Electronics Command (CECOM) broad agency announcement (BAA) (preliminary specifications for the Indigo Systems unit are a D-cell battery, at 0.5 W and weighing 50 g), and Lawrence Livermore National Laboratory is working on a solid-state IR camera for detecting moving objects. When available, these systems should provide small and capable sensors for MUR systems.

Table 2. IR camera specifications.

Specification	VideoTherm 2000	IR microCAM
Detector size (w/o lens)	3.1 w × 3.2 l × 0.9 d	2.75 w × 3.19 l × 0.375 d
Driver size	3.1 w × 4.0 l × 0.47 d	5.5 l × 2.7 w × 1.16 h
Lens length	1.454 in. front of lens to detector	2.25 in. front of lens to detector
Total weight (w/o lens)	5 oz (on postal scale)	7.54 oz
Additional weight	Lenses, mounting fixtures	Lenses, mounting fixtures
Lens weight	0.94 oz	1.27 oz
Spectral range	8–14 μm	8–14 μm
Sensitivity	>0.1 °C thermal resolution	0.07 °C NEDT
Frame rate	30 frames/s	60 frames/s
Lens f-number	f/0.8	f/1
Hyperfocus lens	Yes	Yes
Field of view	50°	33°–25°
Operating temperature range	–20 to + 55 °C	–20 to +60 °C
Settling time	45 s typical	30 s typical
Uniformity method	Rotating mechanical shutter	Block shutter and zero every 2–3 hr (or 1-oz shutter)
Output signal	8 bits/pixel digital	RS-170, 12-bit digital
Blur	Yes, recursive filter, can be removed but image flickers	None (visual check)
Availability	10–12 wk worst case	4 mo
Iris	Yes, manual adjust	No, auto-level
Operating voltage	4.5 to 36 V dc	12 V
Power	Max 6 W, typical is 2–3 W	Max 7.5 W, 4.3 W typical
Camera parts cost	$10,525	$50,000
Lens cost	$1.2 k ($1.6 k with iris)	Included

7. Image Processing for Reconnaissance

The robot's reconnaissance system is required to detect people and traps and report their locations to the operator. A variety of imaging and nonimaging sensors are aboard the robot that can perform these tasks. Each of the robot's sensors has been selected for the synergistic value that it adds to the RSTA system. The visible and IR imaging sensors provide the robot with the sense of vision. With this additional sense, the robot is more capable of detecting and identifying events and objects of interest. The imaging sensors can also locate targets more accurately than the nonimaging sensors. Pictures of significant events and objects can be provided to the operator for positive identification. There will, however, be times when the opposing force can be seen but not heard, or vice versa. Thus, the integration of the imaging and nonimaging sensors will provide superior reconnaissance performance. This section describes the image processing algorithms for person detection using the visible and IR sensors.

There are a number of states of the robot and of objects in the environment that determine the appropriateness of various reconnaissance algorithms. Most important is whether or not the robot and targets of interest are moving. The ability to detect the motion of objects in the environment is extremely important to the survival of virtually all animals. The detection of visual motion will play an equally important role to the microrobot. One of our image-based RSTA algorithms will therefore detect moving objects. When targets in the scene are not moving, they must still be detected. Thus, our other image-based RSTA algorithm will detect people, whether or not they are moving. Because of system constraints as described in the next section, the moving object detection algorithm can only be used when the microrobot is stationary. The person detection algorithm can be used from either a moving or a stationary robot. Table 1 shows which algorithms are applied to which sensors and whether the robot is stationary or moving.

7.1 Moving Object Detection

Moving objects can be detected in image sequences taken from both stationary and moving cameras. Detecting moving objects from a moving camera in an arbitrary environment is challenging, but algorithms have been developed (Irani and Anandan, 1996; Fejes and Davis, 1997) that accomplish this goal. We have observed good results when testing the algorithm of Fejes and Davis (1997) in the robot scenario. Essentially, the algorithm looks for regions of the normal optic flow field (the component of optic flow that can be robustly computed) that fail to satisfy a number of qualitative constraints. Unfortunately, this algorithm is computationally too complex to be implemented in real time on the current robot. Thus, our moving object detection capability is limited to stationary cameras—a much simpler problem.

The moving object detection and tracking algorithm used in this report is based on a real-time system developed for the Office of the Secretary of Defense (OSD) Robotics Demonstration I and II programs (Balakirsky et al, 1993). The system has been extensively tested in indoor and outdoor environments and it performed excellently. In May 1996, the system was successfully deployed for an RSTA mission in an unmanned ground vehicle in field exercises at Ft Hood, TX, by U.S. Army soldiers from the 1st Armored Cavalry Division (David, 1996).

The algorithm performs detection and tracking of moving objects in visible or IR imagery. Through the use of a stationary camera, moving objects can be detected by locating the regions of an image sequence that are changing. Unfortunately, events other than moving objects can cause changes in the image sequence. Events such as sensor noise (which is significant with the IR imager), changing scene illumination, and sensor vibration (caused by, for example, wind or motor vibrations) need to be accounted for. The system first acquires a single-reference image and then compares this with all subsequent images. In the system of David (1996), the reference image was dynamically adapted to represent the static components of the background. In this system, however, because of the expected short duration that a moving target detection is to be applied (a few seconds at a time), a static reference image is adequate. This reference image will not be able to adapt to changing scene illumination, but for such short duration interrogations, this is usually not necessary. (If a sudden change in illumination occurred, such as someone turning on or off the lights, the system can easily detect this event and acquire a new reference image.) As each new frame is acquired, the difference between it and the reference image is computed. The difference image is then thresholded; a single low threshold is applied to the entire image. A binary erosion operation over a 3×3 neighborhood is applied to this binary difference image. This step eliminates many spurious detections because of sensor noise and small camera vibration. Connected regions of these pixels are then grouped into objects described by a number of size and shape properties. A simple target tracker is then used to determine the objects that correspond from one frame to the next. The tracker filters out objects that do not exhibit consistent motion (i.e., clutter). A map of the scene in view, if available, is used to estimate the actual size of the tracked objects, and then simple classification (i.e., animal, person, vehicle) of the objects is performed. Figure 23 illustrates the detection of a person and his reflection in an IR image sequence.

The following briefly describes the average run-time complexity of the moving object detection and tracking algorithm. If we assume that $100R$ percent ($0 \leq R \leq 1$) of the pixels in the difference image exceeds the detection threshold, then the approximate number of integer operations required to execute the above algorithm on a single $N \times N$ image is $N^2 \times (32R + 25) + 900$. The value of R depends on many factors, but for a typical scene of a small number of moving objects, $R = 0.05$ is a good

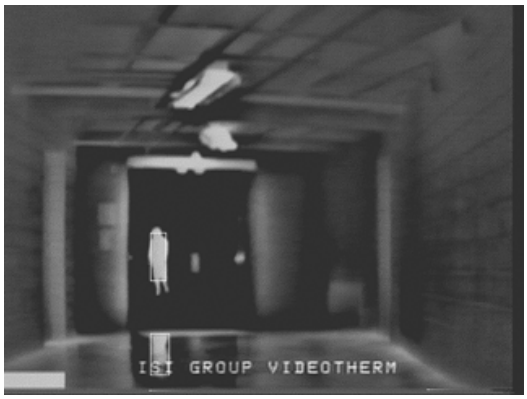


Figure 23. Detection of a moving person and his reflection in an IR image.

empirical estimate of this parameter. Then for an input image of size $N = 512$, the algorithm requires about 6.97 million integer operations. The 180-MHz MIPS R4650 microprocessor on which these algorithms are to run has a stated performance of 90 million integers or 60 million floating point operations a second. Ignoring various system overheads and delays, we can deduce that the above algorithm requires approximately 0.077 s per frame, or equivalently that it can process about 13 frames a second. A processing rate of three to five frames a second is sufficient for the current application. Thus, this algorithm easily fits on the targeted real-time platform. In fact, sufficient CPU time remains for other applications to run in parallel with the motion tracking algorithms. Also, there may be times when it is desirable to perform moving object tracking on multiple imaging sensors simultaneously. This is easily implemented with these algorithms, up to four cameras at a time (about three frames a second per camera).

7.2 Person Detection

A requirement of the robot's RSTA system is to detect people in the environment. In many situations, the robot's imaging sensors provide enough information to accomplish this goal. The moving object detection and tracking system described in the previous section reliably detects people who are moving in either visible or IR imagery. What if the people are not moving? Then that algorithm fails to detect them. A different algorithm is required that detects stationary people. Since the people are assumed to be stationary, such an algorithm should require only a single image to perform this function. The algorithm should handle a wide variety of poses of people standing, sitting, kneeling, lying, and facing toward and away from the camera, etc, and it should work for partially occluded people positioned at a variety of ranges.

The computer vision community has pursued a variety of approaches to person detection that might be applicable to this type of robotic platform. One that has received significant attention is human face detection. While the face is probably the most distinctive part of the human body, most

faces are still geometrically quite similar. This similarity allows algorithms to perform detection via a search for a generic face and then perform recognition, if necessary, based on particular features of the face. Some of the techniques used to detect faces include neural nets (Lin et al, 1997; Rowley et al, 1998), eigenfaces (Facia Reco Associates, 1997), and geometric model matching (Jeng et al, 1998). All these algorithms attempt to locate the faces of people who are looking approximately toward the camera. Most of these algorithms fail for off-angle faces, and all fail for faces that are not visible, such as for a person walking away from the camera. Still, we tested a number of face detection algorithms that were designed to work with visible imagery and observed moderate to poor performance in uncontrolled, cluttered environments: numerous faces went undetected, many false detections were generated, and at times, the programs were very slow. There are head detection algorithms (Sirohey, 1993; Birchfield, 1997) that attempt to locate the elliptically shaped human head and thus may not suffer as many of the problems as face detection algorithms. However, both face and head detection algorithms have difficulties detecting people in cluttered environments, especially if the image of the head is small enough that it loses some of its distinguishing features. Some other approaches rely on color models of human skin to detect people (Oliver et al, 1996; Fieguth et al, 1997). These approaches will not work well in the dimly lit (or dark) environments in which the robot is expected to operate. Oren et al (1997) have developed a wavelet-based approach for detecting the full bodies of pedestrians moving toward or away from a camera. This approach, too, has limited use for our application. In general, image-based detection of unconstrained people in unconstrained environments is an extremely difficult task that requires further research for reliable solutions to be developed. Maybe some combination of the above techniques might prove fruitful.

After considering the difficulty of detecting stationary people and the limited computational resources available in our robot, we decided to investigate a simple technique based on detecting person-shaped blobs in IR imagery. Although the technique was not expected to solve the problem, its performance is sufficient to provide useful information in many situations. Entire (or partial) bodies are detected rather than just heads and faces. This results in a more reliable system when the people are distant from the camera. IR rather than visible imagery is used because people are easily detected in IR imagery as high-intensity blobs (i.e., hot spots) by simple thresholding techniques. Hot objects in the scene other than people may also be detected by this simple method, so a means to discriminate between person-like and non-person-like blobs is necessary.

The steps of the algorithm follow. A single relatively low threshold is applied uniformly to the entire image. This results in a binary image of blobs. To smooth the boundaries of the blobs to improve the quality of the image, shape analysis, erosion, and then dilation are applied to this

binary image. This step can also remove many small blobs caused by noise and small clutter. Connected regions of these pixels are then grouped into objects with the use of a standard eight-connected component algorithm. Next, the perimeter of each blob is traced. Multiple straight lines are fitted around the perimeter of each blob as it is being traced. A particular line ends and a new line is started when the current pixel on the perimeter is greater than some distance (typically one to three pixels) from the current line (the ratio of the number of high-contrast perimeter pixels to the total number of perimeter pixels). A perimeter pixel is deemed high-contrast if the intensity gradient across the edge exceeds some fixed threshold. This is the same contrast measure as that used by Birchfield (1997) for human head detection via elliptical contour fitting.

The above process results in two parameters for each blob: the number of lines fitted to its perimeter and the perimeter contrast ratio. Simple pattern classification techniques are then used to classify the blob as either person or nonperson based on these two parameters. The number-of-lines parameter describes the shape of the blob. In general, we have found that the IR signatures of most inanimate objects found in indoor environments are blobs whose shapes can be approximated by a small number of straight lines. These are typically objects such as windows, doorways, corners of hallways, and lights. Sometimes, however, when the intensity of an object's image is close to the threshold used to generate the binary blob image, blobs with contorted shapes are generated because of small fluctuations in temperature across the surface of the object. In this case, the contrast across the part of the blob's perimeter that is a product of these temperature fluctuations will usually be small, and hence the blob's contrast ratio parameter should also be small. The shapes of people, on the other hand, are usually not well approximated by a small number of straight lines, and their contrast with the background is usually quite good. Thus, we expect a good separation of people from nonpeople in the 2-D parameter space that we have developed.

These ideas have been tested on a variety of IR imagery of people in different environments, in different poses, and at different ranges. Our experiments consisted of running approximately 200 IR images through the above analysis. The blobs generated by each image were manually classified as either person or nonperson. The parameters of each were then plotted in a scatter diagram to determine an appropriate discrimination function. This is illustrated in figure 24. From this scatter diagram, it is obvious that our metrics do not allow linear separation of people from nonpeople. However, the piecewise linear discriminant function shown (by the two dashed lines) allows for a reasonably good separation. (In the current system, this discriminant function is determined manually.) For the data analyzed so far, they result in a 0.97 probability of correctly classifying person blobs as people (probability of detection) and a

0.02 probability of incorrectly classifying a nonperson blob as a person (probability of false alarm). Figure 25 illustrates the results of processing a typical frame of IR imagery. Depending on various assumptions made about the imagery being processed, this algorithm produces about one false alarm for every ten or so images analyzed. By itself, this false alarm rate is still too high for practical use as a person detector. But, when the person-detection algorithm is used with other algorithms and sensors, this algorithm should be quite helpful in separating people from nonpeople.

The average run-time complexity of this algorithm is now briefly described. If we assume that $100S$ percent ($0 \le S \le 1$) of the pixels in the image exceeds the detection threshold and assume 10 blobs per frame with an average perimeter length of 240 pixels, then the approximate number of operations required to execute the above algorithm on a single $N \times N$ image is $N^2 \times (180S + 23) + 120{,}400$ integer operations plus an additional 0.84 million floating point operations. The value of S depends on many factors. But for a typical scene, $S = 0.07$ is a reasonable estimate. Then, for an input image of size $N = 512$, the algorithm requires about 9.5 million integers plus 0.84 million floating point operations. Ignoring various system overheads and delays implies that the above algorithm executes in approximately 0.11 s per image. Thus, this algorithm too easily fits on the targeted real-time platform.

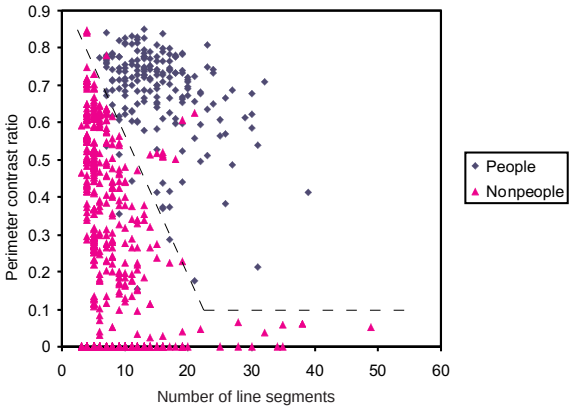


Figure 24. Scatter diagram of blob perimeter contrast ratio vs number of lines required to approximate the blob's perimeter.



Figure 25. Discrimination of people from other "hot spots" in IR imagery.

8. Future Directions

The first major direction that will be pursued in the future is to extend the system from a single urban agent to multiple cooperative urban agents that use a single operator interface or a mother ship. Multiple urban agents that operate collaboratively can (1) act as sentries for one another, (2) extend communications by acting as relays when going deep inside a building, and (3) act as teams to clear buildings more effectively and efficiently.

The addition of a mother ship and its processing capabilities will allow the system to take advantage of ongoing visualization research at ARL that can be used to enhance the mapping of terrain; weather; nuclear, biological, and chemical (NBC) environments; and the interiors of the buildings that the rovers are exploring. These additional computational capabilities can also greatly enhance the mission planning, autonomous navigation, and collaborative effort of the system.

For missions that are predominantly urban, the small rovers discussed in this report are the anticipated platforms because of their more agile mobility characteristics. When the mission turns to open terrain in the field, we project that larger rovers with longer ranges and payloads will be better suited. The mother ship could then deliver either type of rover and the appropriate RSTA module or a combination of both, depending on the terrain and mission. An important area of interest with the mother-ship concept is to develop a docking system for the rovers to dock to the mother ship. In addition to the physical and electrical connections that must be made, the mother ship must be able to communicate with the rovers and the rovers with each other. Therefore, several communication experiments are planned to test the appropriate communication modes.

Another direction that could be taken is to use the agents in a logistical role. The agents can be used to replace batteries, exchange sensors, retrieve robots, and repair damaged components. Even entire modules can be replaced. Done autonomously, this logistics agent can greatly extend the range and life of the rovers as the missions or scenarios change during an operation.

Another direction that is of interest is to extend the agents' environment to the littoral battlespace. By developing sensors to work in the surf zone and rovers that can transition from the sea to land, stealth, surprise, and deeper penetration into enemy territory can be obtained and delivered from a farther distance with torpedo-based capsules to deliver the rovers.

The military applications that we are primarily developing for the rovers can be easily extended to operations other than war. The rovers can be used by civilian authorities to search buildings for victims after a disaster or to search for missing children in the woods and to find chemical spills quickly and without risk to humans in a city or in the country.

9. Conclusion

The application of small robots in detecting hazards in urban warfare scenarios is indeed feasible. The strong trend of sensor miniaturization and processor efficiency supports this concept. The initial system analysis and laboratory experimentation of the individual sensors have been highly supportive of this conclusion. The simple multisensor fusion approach that is planned is very low risk. The most difficult tasks remaining are to scale and map the image-detection algorithms to the vehicle processor. However, because of the feasibility of using small robots for this application, the multiagent collaborative RSTA planning research has been initiated.

References

Balakirsky, Stephen, Phil David, Philip Emmerman, Francis Fisher, and Phil Gaylord (1993) "Semi-Autonomous Mobile Target Engagement System," Symp. Association for Unmanned Vehicle Systems (June).

Birchfield, Stephen (1997), "An Elliptical Head Tracker," *Proc. of the 31st Asilomar Conf. Signals, Syst., and Comput.* (November).

David, Phil (1996), "Real-Time Target Acquisition for the Army's Scout Mission," *Proc. SPIE, Acquis., Tracking, and Pointing X*, **2739** (April).

Drago, P. G., et al (1978), "Digital Dynamic Speech Detectors," *IEEE Trans. Commun.*, **Com-26**, No. 1 (January).

Facia Reco Associates (1997), developed commercial software, Waltham, MA.

Fejes Sandor and Larry Davis (1997), *Detection of Independent Motion Using Directional Motion Estimation*, University of Maryland at College Park CS-TR-3814, CAR-TR-865 (July).

Fieguth P., and D. Terzopoulos (1997), "Color-Based Tracking of Heads and Other Mobile Objects at Video Frame Rates," *Proc. IEEE Conf. Comput. Vision and Pattern Recognition*, pp 21–27.

Irani, M., and P. Anandan (1996), "A Unified Approach to Moving Object Detection in 2D and 3D Scenes," *Proc. ARPA Image Understanding Workshop*, Palm Springs, CA, pp 707–718 (February).

Jeng, Shi-Hong, Hong Liao, Chin Han, Ming Chern, and Yao Liu (1998), "Facial Feature Detection Using Geometrical Face Model: An Efficient Approach," *Pattern Recognition*, **31**, No. 3, pp 273–282.

Lin, S, S. Kung, and L. Lin (1997), "Face Recognition/Detection by Probabilistic Decision-Based Neural Network," *IEEE Trans. Neural Networks*, **8**, No. 1, pp 114–132 (January).

Makino, S., and K. Kido (1996), "Recognition of Phonemes Using Time-Spectrum Pattern," *Speech Commun.*, **5**, No. 2 (June).

Newman, William C. (1990), "Detecting Speech With An Adaptive Neural Network," *Electron. Des.* (22 March).

Oren, M., C. Papageorgiou, P. Sinha, E. Osuna, and T. Poggio (1997), "A Trainable System for People Detection," *Proc. DARPA Image Understanding Workshop*, New Orleans, LA, pp 207–214.

Oliver, N., S. Pentland, F. Berard, and J. Coutaz (1996), *LAFTER: Lips and Face Tracker, MIT Media Laboratory Perceptual Computing Section*, Technical Report 396.

Rowley, H., S. Baluja, and T. Kanade (1998), "Neural Network-Based Face Detection," *IEEE Trans. Pattern Anal. and Mach. Intelligence*, **20**, No. 1 (January).

Sirohey, Sirohey (1993), *Human Face Segmentation and Identification*, University of Maryland at College Park, CS-TR-3176, CAR-TR-695 (November).

Taboada, J., et al (1994) ,"Explicit Estimation of Speech Boundaries," *IEEE Proc. Sci. Meas. Technol.*, **141**, No. 3 (May).

Turk, M., and A. Pentland (1991), "Eigenfaces for Recognition," *J. Cognitive Neuros.*, **3**, No. 1, pp 71–86.

Waibel, A., et al (1989), "Phoneme Recognition Using Time-Delay Neural Networks," *IEEE Trans. Acoust., Speech, Signal Process.*, **37**, No. 3 (March).

Distribution

Admnstr
Defns Techl Info Ctr
Attn DTIC-OCP
8725 John J Kingman Rd Ste 0944
FT Belvoir VA 22060-6218

Ofc of the Dir Rsrch and Engrg
Attn R Menz
Pentagon Rm 3E1089
Washington DC 20301-3080

Ofc of the Secy of Defns
Attn ODDRE (R&AT)
The Pentagon
Washington DC 20301-3080

OSD
Attn OUSD(A&T)/ODDR&E(R) R J Trew
Washington DC 20301-7100

AMCOM MRDEC
Attn AMSMI-RD W C McCorkle
Redstone Arsenal AL 35898-5240

CECOM
Attn PM GPS COL S Young
FT Monmouth NJ 07703

Dir for MANPRINT
Ofc of the Deputy Chief of Staff for Prsnnl
Attn J Hiller
The Pentagon Rm 2C733
Washington DC 20301-0300

US Army Armament Rsrch Dev & Engrg Ctr
Attn AMSTA-AR-TD M Fisette
Bldg 1
Picatinny Arsenal NJ 07806-5000

US Army Edgewood RDEC
Attn SCBRD-TD G Resnick
Aberdeen Proving Ground MD 21010-5423

US Army Info Sys Engrg Cmnd
Attn ASQB-OTD F Jenia
FT Huachuca AZ 85613-5300

US Army Natick RDEC Acting Techl Dir
Attn SSCNC-T P Brandler
Natick MA 01760-5002

Director
US Army Rsrch Ofc
4300 S Miami Blvd
Research Triangle Park NC 27709

US Army Simulation, Train, & Instrmntn Cmnd
Attn J Stahl
12350 Research Parkway
Orlando FL 32826-3726

US Army Tank-Automtv Cmnd Rsrch, Dev, & Engrg Ctr
Attn AMSTA-TA J Chapin
Warren MI 48397-5000

US Army Train & Doctrine Cmnd
Battle Lab Integration & Techl Dirctrt
Attn ATCD-B J A Klevecz
FT Monroe VA 23651-5850

US Military Academy
Mathematical Sci Ctr of Excellence
Attn MDN-A MAJ M D Phillips
Dept of Mathematical Sci Thayer Hall
West Point NY 10996-1786

Nav Surface Warfare Ctr
Attn Code B07 J Pennella
17320 Dahlgren Rd Bldg 1470 Rm 1101
Dahlgren VA 22448-5100

DARPA
Attn B Kaspar
3701 N Fairfax Dr
Arlington VA 22203-1714

Hicks & Associates, Inc
Attn G Singley III
1710 Goodrich Dr Ste 1300
McLean VA 22102

US Army Rsrch Lab
Attn AMSRL-DD J Rocchio
Attn AMSRL-CI-LL Techl Lib (3 copies)
Attn AMSRL-CS-AS Mail & Records Mgmt
Attn AMSRL-CS-EA-TP Techl Pub (3 copies)
Attn AMSRL-IS J D Gantt
Attn AMSRL-IS P Emmerman (3 copies)

Distribution (cont'd)

US Army Rsrch Lab (cont'd)
Attn AMSRL-IS-C COL M R Kindl
Attn AMSRL-IS-CB J Locker
Attn AMSRL-IS-CB J Vanetten
Attn AMSRL-IS-CB L Tokarcik
Attn AMSRL-IS-CB P Budulas (3 copies)
Attn AMSRL-IS-CB R Gregory
Attn AMSRL-IS-CB S Bragonier
Attn AMSRL-IS-CB T Rose
Attn AMSRL-IS-CI B Broome
Attn AMSRL-IS-CI D Hillis
Attn AMSRL-IS-CB P David

US Army Rsrch Lab (cont'd)
Attn AMSRL-IS-CI P Fisher
Attn AMSRL-IS-CB S Young (3 copies)
Attn AMSRL-IS-CI T Mills
Attn AMSRL-IS-E K Gurton
Attn AMSRL-IS-ES S Crow
Attn AMSRL-SE-EA M Scanlon
Attn AMSRL-SE-RU A Harrison
Attn AMSRL-SE-SA J Eicke
Attn AMSRL-SE-SA M Wellman
Attn AMSRL-SE-SA N Srour
Attn AMSRL-SE-SS S Vinci
Adelphi MD 20783-1197

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	May 1999	Final, 8/1997–3/1998

4. TITLE AND SUBTITLE RSTA for Small Rovers in Urban Warfare

5. FUNDING NUMBERS

DA PR: —
PE: —

6. AUTHOR(S) Stuart Young, Pete Budulas, Philip Emmerman, Mike Scanlon, Nassy Srour, Dave Hillis, Phil David, Pete Fisher, Steve Vinci, Art Harrison, Kris Gurton, Sam Crow, and Mark Wellman

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Army Research Laboratory
Attn: AMSRL-IS-CB email: syoung@arl.mil
2800 Powder Mill Road
Adelphi, MD 20783-1197

8. PERFORMING ORGANIZATION REPORT NUMBER
ARL-TR-1678

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
DARPA
3701 N Fairfax Dr
Arlington VA 22203-1714

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
ARL PR: 78LAZ4
AMS code: P63709D

12a. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The U.S. Army Research Laboratory (ARL) is pursuing a major research initiative in robotics. This research centers on collaborative physical agents that have advanced sensing, analysis, and behavioral characteristics that are linked to a mother ship that uses advanced visualization and awareness tools. Applications that are prompting this effort are reconnaissance, surveillance, and target acquisition (RSTA) of both human and vehicle targets as well as nuclear, biological, and chemical agent detection and localization. This report focuses on the requirements of a robot, or rover, to operate in urban terrain (such as military operations in urbanized terrain (MOUT) facility), to autonomously and stealthily approach enemy-controlled buildings, and to identify humans and any hazards to them. The requirements for this scenario could be performed by three increasingly complex systems, depending upon the extent of the operation. The three proposed systems are an individual agent, a team of collaborative agents, or a mother ship that works with a team of collaborative agents. This report focuses on a single physical agent solution. The agents must be able to negotiate all areas, such as curbs, stairs, and rubble, within an urban terrain. This report also discusses the application and component research thrusts of the RSTA module to detect humans and hazards to humans.

14. SUBJECT TERMS
Physical agents, target acquisition, mother ship

15. NUMBER OF PAGES
50

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	SAR

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

DEPARTMENT OF THE ARMY
U.S. Army Research Laboratory
2800 Powder Mill Road
Adelphi, MD 20783-1197

An Equal Opportunity Employer