Unmanned ground vehicle perception using thermal infrared cameras

Arturo Rankin*, Andres Huertas, Larry Matthies, Max Bajracharya, Christopher Assad,
Shane Brennan, and Paolo Bellutta
Jet Propulsion Laboratory, California Institute of Technology
4800 Oak Grove Drive, Pasadena, CA, USA 91109

Gary W. Sherwin
General Dynamics Robotic Systems
1501 Ardmore Boulevard, Pittsburgh, PA, USA 15221

ABSTRACT

The ability to perform off-road autonomous navigation at any time of day or night is a requirement for some unmanned ground vehicle (UGV) programs. Because there are times when it is desirable for military UGVs to operate without emitting strong, detectable electromagnetic signals, a passive only terrain perception mode of operation is also often a requirement. Thermal infrared (TIR) cameras can be used to provide day and night passive terrain perception. TIR cameras have a detector sensitive to either mid-wave infrared (MWIR) radiation (3-5μm) or long-wave infrared (LWIR) radiation (7-14μm). With the recent emergence of high-quality uncooled LWIR cameras, TIR cameras have become viable passive perception options for some UGV programs. The Jet Propulsion Laboratory (JPL) has used a stereo pair of TIR cameras under several UGV programs to perform stereo ranging, terrain mapping, tree-trunk detection, pedestrian detection, negative obstacle detection, and water detection based on object reflections. In addition, we have evaluated stereo range data at a variety of UGV speeds, evaluated dual-band TIR classification of soil, vegetation, and rock terrain types, analyzed 24 hour water and 12 hour mud TIR imagery, and analyzed TIR imagery for hazard detection through smoke. Since TIR cameras do not currently provide the resolution available from megapixel color cameras, a UGV's daytime safe speed is often reduced when using TIR instead of color cameras. In this paper, we summarize the UGV terrain perception work JPL has performed with TIR cameras over the last decade and describe a calibration target developed by General Dynamics Robotic Systems (GDRS) for TIR cameras and other sensors.

Keywords: Thermal infrared, MWIR, LWIR, stereo vision, passive perception, terrain classification, calibration

1. INTRODUCTION

Active and passive sensors have been successfully used to accomplish unmanned ground vehicle (UGV) autonomous navigation. Active sensors emit radiant energy to illuminate a scene while passive imaging systems capture existing scene illumination. The passive imaging options for daytime UGV operation are color, near-infrared (NIR), short-wave infrared (SWIR), cooled forward looking infrared (FLIR), and uncooled FLIR cameras. The passive imaging options for nighttime UGV operation are image intensifiers, cooled FLIR cameras, and uncooled FLIR cameras. Image intensifiers amplify the existing scene illumination by electronic means. Although image intensifiers are lower cost, FLIR cameras are preferred because they have a much higher signal to noise (SNR) ratio. The SNR of image intensifiers can be improved by increasing the exposure time, but this can lead to significant image blur problems during UGV motion. A FLIR camera is the only passive option that provides useful imagery over the entire 24-hour cycle.

A 1999 study determined that cooled FLIR cameras were the only passive imaging option that would enable nighttime autonomous navigation at speeds up to 4.5m/s [1]. In the last decade, however, the image quality and resolution of uncooled FLIR cameras have significantly improved. The primary technical metric to consider when choosing between currently available cooled and uncooled FLIR cameras is the exposure time necessary to achieve required UGV speeds. High levels of image blur will degrade the performance of stereo ranging, terrain classification, and feature identification and tracking. For this reason, cooled FLIR systems are preferred for high speed UGVs. However, cost is usually also a major factor in choosing between a cooled and uncooled FLIR. There are two basic types of FLIR systems; mid-wave infrared (MWIR) cameras, which are sensitive to 3-5μm radiant energy, and long-wave infrared (LWIR) cameras, which

*Arturo.Rankin@jpl.nasa.gov; phone (818) 354-9269; fax (818) 393-5007; http://robotics.jpl.nasa.gov

are sensitive to 7-14μm radiant energy. Since MWIR and LWIR cameras operate within the thermal infrared (TIR) portion of the electromagnetic spectrum, both types are considered a TIR camera. Both cooled MWIR and LWIR cameras have adequate thermal measurement sensitivity to provide adequate contrast for stereo matching and autonomous navigation under varying thermal conditions and degraded atmospheric transmission situations. A stereo pair of TIR cameras has been used to generate sparse stereo disparity data when the cameras are uncalibrated [2] and dense stereo range data when the cameras are calibrated [3].

Some military UGV programs require an ability to perform autonomous navigation during the day and night, and an ability to do it in a stealth mode (where perception sensor energy is not emitted). TIR cameras can be used to accomplish both requirements. If a pair of TIR cameras is used, stereo ranging and terrain classification can be performed to generate an annotated map of the terrain. TIR is a convenient option since a single TIR camera may already be a part of the sensor suite of many vehicles. Military UGVs are likely to carry TIR cameras to aid in perception at night and through atmospheric obscurants, and commercial passenger cars are also now available with TIR cameras to aid drivers during night operation.

Over the last decade, JPL has developed a variety of terrain perception algorithms applicable to a stereo pair of TIR cameras. Under the Army Research Laboratory (ARL) Demo III program [as a subcontractor to General Dynamics Robotic Systems (GDRS)], JPL developed and field tested positive obstacle, negative obstacle, and tree-line detection algorithms using a stereo pair of Indigo Systems Merlin cooled MWIR cameras [3][4]. In addition, JPL and GDRS collected stereo Merlin MWIR imagery in fog-oil. Under the DARPA Perception for Off-Road Robots (PerceptOR) program [as a subcontractor to Science Applications International Corporation (SAIC)], JPL performed 24-hour cross-country data collections and developed and field tested terrain classification, negative obstacle detection, and tree-trunk detection algorithms using a stereo pair of Cincinnati Electronics NC256 cooled MWIR cameras. JPL also tested a dual TIR band classifier using a NC256 MWIR camera and an Indigo Systems Alpha uncooled LWIR camera [5][6][7][8]. In addition, JPL and SAIC collected NC256 MWIR stereo imagery in HC smoke. Under the ARL Robotics Collaborative Technology Alliances (RCTA) program (as a consortium partner with GDRS and others), JPL developed pedestrian, vehicle, and water detection algorithms [9][10][11][12][14]. These algorithms are currently being tested on imagery collected from a stereo pairs of FLIR Systems Photon 640 uncooled LWIR cameras and Thermoteknix Miricle 110KS uncooled LWIR cameras. In addition, GDRS has collected stereo imagery thru fog and smoke using a stereo pair of Photon 640 LWIR cameras. This paper summarizes the UGV terrain perception work JPL has performed with TIR cameras under these programs and describes a calibration target developed by GDRS for calibrating both color and TIR cameras.

2. STEREO RANGING USING TIR CAMERAS

2.1 Stereo ranging during UGV motion

During Demo III and PerceptOR, JPL used a stereo pair of Merlin and NC256 cooled MWIR cameras to perform passive terrain perception at night. At that time, the image quality and relatively low image resolution of available uncooled LWIR cameras made them a poor option for stereo ranging. In recent years, however, the image quality and resolution of uncooled LWIR cameras have significantly improved, making them a viable option for UGVs that operate at slow speed. Fig. 1 contains sample images from cooled MWIR and uncooled LWIR cameras from a decade ago, and an uncooled LWIR image acquired last year with a newer LWIR camera. One can clearly see the older uncooled LWIR camera has a much lower SNR.

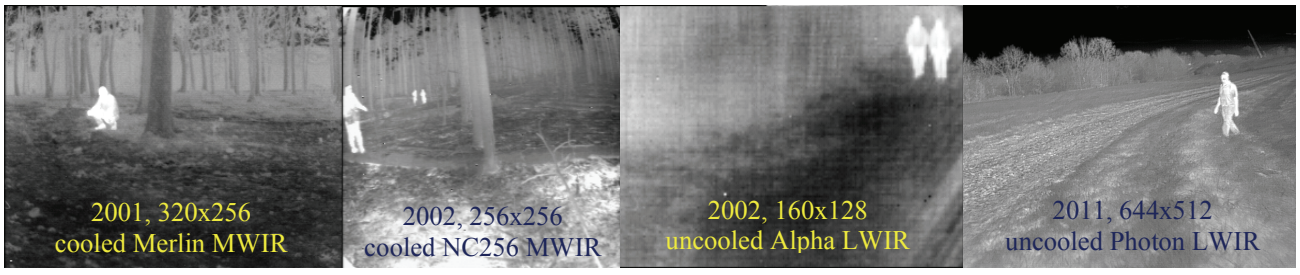


Fig. 1. The quality and resolution of uncooled LWIR cameras have increased significantly over the last decade, rivaling cooled TIR cameras that were available a decade ago.

Table 1 contains the key specifications for the commercial-off-the-shelf (COTS) TIR cameras illustrated in this paper. From Table 1, the Photon 640 uncooled LWIR camera has a smaller pixel pitch, weight, size, and power requirement, and a higher resolution than the Merlin and NC256 cooled MWIR cameras that were the state-of-the-art a decade ago. However, the sensitivity of the Photon 640 is twice that of the Merlin and NC256. The sensitivity of uncooled LWIR cameras continues to improve though. There is a version of the Miricle 110KS LWIR camera available now that provides a sensitivity of 10-35 mK noise equivalent differential temperature (NEdT), which rivals the Merlin and NC256. Today, the primary concern regarding using uncooled LWIR cameras for UGV terrain perception is their exposure time. Even here there have been significant gains over the last decade. Exposure times have decreased from 33ms (Indigo Systems Alpha) to 7ms (Thermoteknix Miricle 110KS). In this section, we focus on the applicability of currently available uncooled LWIR cameras for real-time UGV stereo ranging.

Table 1. Key specifications for the COTS TIR cameras illustrated in this paper.

Manufacturer Model	**Camera Description**	**Spectral Sensitivity (nm)**	**Integration time**	**NEdT (mK)**	**Power (W)**	**Dimensions H x W x L (cm) Mass (g) *without lens**	**Resolution Dyn. range Pixel pitch Focal length Field of view**
Indigo Systems Merlin	MWIR camera with cooled InSb sensor	3000-5000	5µs – 16.5ms	<25	30	14x12.7x24.9* 4082*	320x256 12bit 30µm 13mm 41°x31°
Cincinnati Electronics NC256	MWIR camera with cooled InSb sensor	3600-5000	100µs – 16.5ms	<20 (w/o lens)	16	7.1x7.1x29.5* 1350*	256x256 14bit 30µm 7mm 67.5°x62.4°
Indigo Systems Alpha	uncooled LWIR camera with micro-bolometer	7500-13,500	33ms	Not known	1.5	4.3x4.3x10.9 <200	160x128 12bit 51µm 11mm 40x30
Thermoteknix Miricle 110KS	uncooled LWIR camera with Alpha Silicon micro-bolometer	7000-14,000	7ms	<=50 @ 30°C	<3.3	4.8x4.3x5.6* 140*	384x288 14bit 35µm 15mm 48°x37°
FLIR Systems Photon 640	uncooled LWIR camera with VoX micro-bolometer	7,500-13,500	13-14ms	<50	<3	6.2x6.4 x 4.6* 170*	644x512 14bit 25µm 21.5mm 41°x33°

The primary causes of loss of stereo data when using TIR cameras is 1) image blur from camera horizontal, vertical, and rotational motion, 2) the presence of low texture regions in a scene, and 3) poor camera calibration. During RCTA, JPL and GDRS imaged cross-country scenes with uncooled LWIR cameras from a stationary and moving UGV. Fig. 2 illustrates the dense stereo range images possible with uncooled LWIR cameras for scenes imaged from a stationary UGV. During UGV motion, if the motion of an uncooled LWIR camera is too large, image blur will occur, reducing the quality of stereo range data. For adequate stereo performance, pixel motion should be limited to less than ½ a pixel over the exposure time. Fig. 3 illustrates the geometry of calculating image blur for a point on the ground, assuming the ground surface is level. The pixel motion P can be calculated for a point a distance L along the ground given a camera's height H, exposure time T_{exp}, instantaneous vertical field of view ($IVFOV$), forward velocity ($v_{forward}$), vertical velocity (v_{up}), and pitch rate ω (1).

Fig. 4 shows a sample 324x256 Photon 320 LWIR image that contains significant image blur, primarily due to changes in the camera's pitch while driving over bumpy terrain. With a time constant of 13-14ms, the pitch rate of the higher resolution Photon 640 camera would need to be limited to a maximum of 2.5°/s at full image resolution and 5.0°/s at half image resolution to avoid image blur greater than ½ a pixel. With a time constant of 7ms, the pitch rate of the Miricle 110KS camera would need to be limited to a maximum of 9.2°/s to avoid image blur greater than ½ a pixel at full image resolution. Fig. 4 also contains a plot of the expected image blur for a Miricle 110KS camera mounted 1.53 meters above a level ground surface with a down tilt of 5°, for UGV speeds of 4, 6, and 8m/s. The curves start where the bottom edge of the camera's field of view (FOV) intersects the ground. Note that for horizontal camera motion, the blur in the image increases as the distance along the ground decreases. For UGV speeds of 4, 6, and 8m/s (Fig. 5), the expected image blur is below ½ a pixel at distances along the ground greater than 6.0, 7.3, and 8.6 meters, respectively.

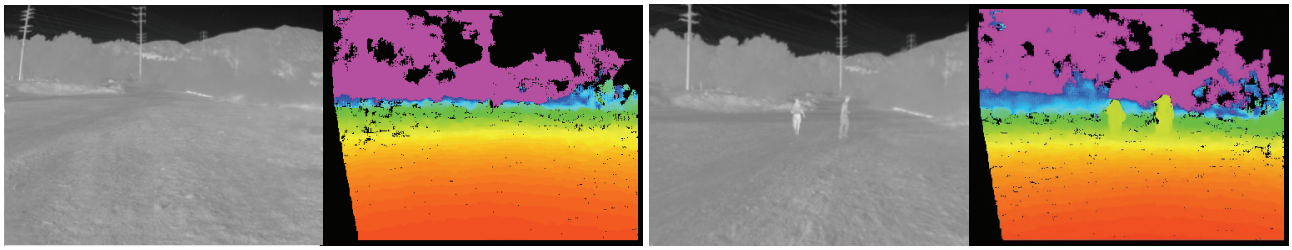


Fig. 2. Sample stereo range images of cross-country terrain generated during a sunny day (at 3:00pm) with a pair of Miricle 110KS cameras mounted to a stationary UGV with a 31cm baseline. Automatic gain control (AGC) was performed over the entire image. The range images are very dense.

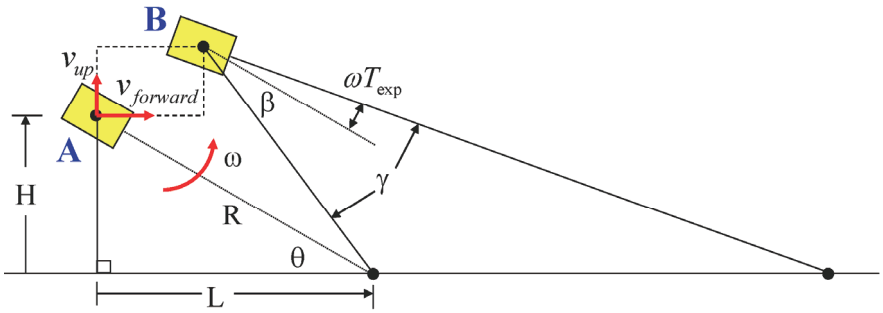
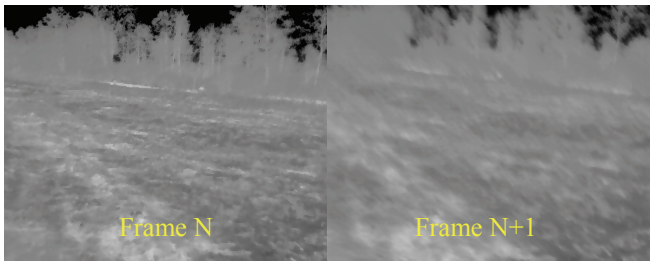
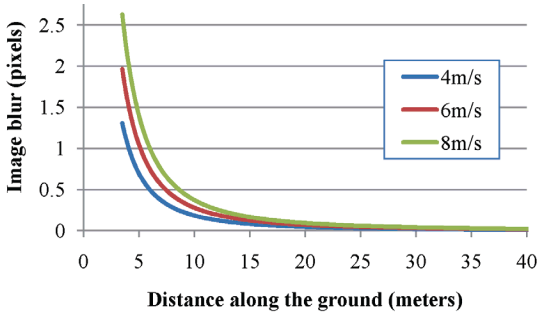


Fig. 3. The geometry of image blur for a point on a level surface. γ represents the vertical angular motion in a an image of a point L meters down range as a result of forward, vertical, and rotational camera motion during the exposure time.

$$P = \frac{\tan^{-1}\left[\dfrac{v_{forward}\,T_{exp}\sin\theta + v_{up}\,T_{exp}\cos\theta}{R-\left(v_{forward}\,T_{exp}\cos\theta - v_{up}\,T_{exp}\sin\theta\right)}\right] + \omega T_{exp}}{IVFOV} \qquad (1)$$



Photon 320: Image blur primarily from high pitch rate



Miricle 110KS: Expected image blur on level ground

Fig. 4. Rotational camera motion (left) and translational camera motion (right) contribute to image blur.

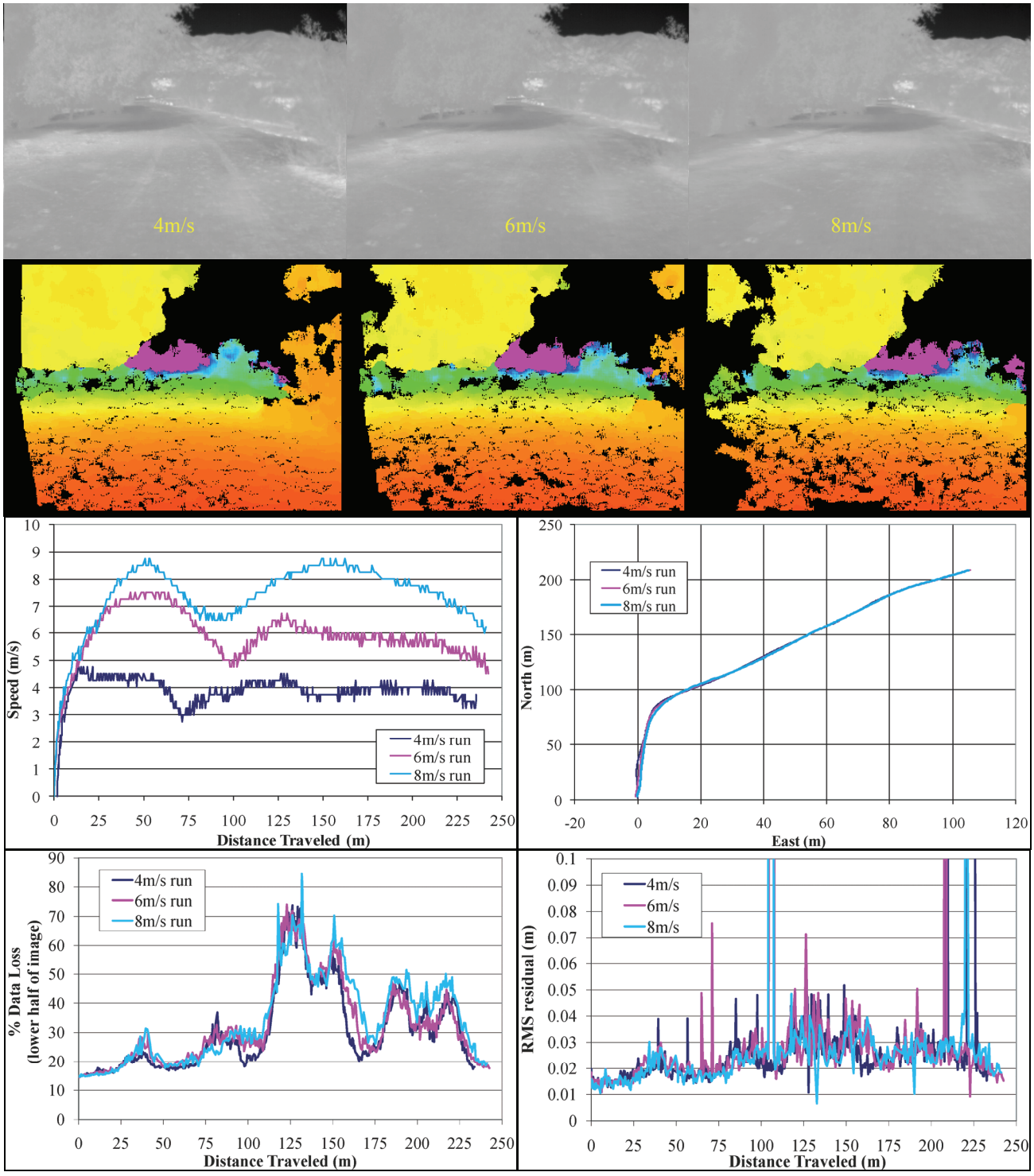


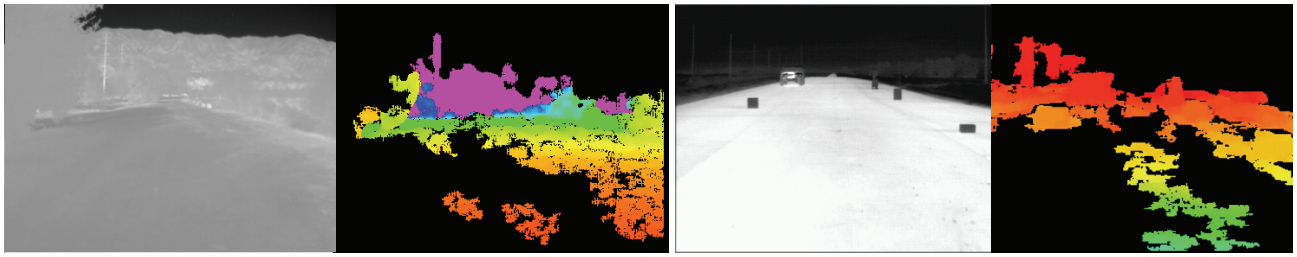
Fig. 5. A UGV was manually driven over the same path at 4, 6, and 8 m/s and stereo pairs of Miricle 110KS LWIR images were acquired and processed off-line. The camera baseline was 31cm. The upper two rows contain intensity and stereo range images for a frame from each run near the same location. Plots of the UGV speed, course, percent stereo data loss in the lower half of the images, and the least squares plane fit RMS residual of a 1 square meter patch of terrain 5 meters in front of the UGV are included. AGC was performed over the entire image.

To experimentally test the quality of range data generated from a stereo pair of uncooled LWIR cameras during camera horizontal motion, a UGV was manually driven over a dirt trail at 4, 6, and 8 m/s and Miricle 110KS images were acquired and processed off-line. The vehicle pose was saved in the header of each LWIR image using a Honeywell Modular Azimuth Positioning System (MAPS) Inertial Navigation System (INS). Fig. 5 contains intensity and stereo range images for a frame from each sequence near the same location. For each sequence, Fig. 5 also contains a plot of the UGV speed, route, data loss in the lower half of the image, and least squares plane fit RMS residual for a 1 square meter patch of terrain virtually positioned at a constant distance of 5 meters in front of the UGV. In the RMS residual plot, the values are mostly below 5cm with the amplitude of the three curves being very similar. The largest spikes are outliers that were caused by the 1 square meter terrain patch spanning objects on the sides of the road.

In the three stereo range images, there is progressively less data (albeit minor) in the lower half of the image with increasing speed. But note that there are portions of the course where the data loss is significant at all speeds. The major factor causing data loss at these speeds (on relatively smooth terrain) was low texture in the scene, not image blur. The left side of Fig. 6 contains an intensity image and stereo range image for a 4m/s frame where there was significant data loss. In comparison to the intensity images in Fig. 5, there is clearly less texture on this portion of the course.

Lack of texture where there is fairly uniform scene temperature is problematic for cooled MWIR cameras as well. The right side of Fig. 6 contains an intensity image and stereo range image for a scene containing significant image saturation on a hot blacktop road. One contributing factor is dynamic range is lost when TIR imagery is scaled to 8 bits per pixel for JPL stereo processing. The problem is exacerbated when there are large temperature extremes in the FOV (such as the sky and pavement on a sunny day) and AGC is performed over the entire scene. One way to reduce the data loss is to perform AGC on a region of the scene that excludes the sky, effectively increasing the dynamic range of low texture regions. Fig. 7 contains range images for frames acquired from approximately the same location during two 12m/s runs down a dirt trail. During the first run, AGC was performed over the entire image. During the second run, when AGC was performed over the lower 3/4th of the image, there was less data loss along the wheel tracks on the trail.

The third cause of TIR stereo range data loss is a poor camera calibration. The next section discusses the technique GDRS uses to calibrate a stereo pair of TIR cameras and the calibration target they designed for use with both color and TIR cameras.



Miricle 110KS LWIR: Low texture section of a dirt trail Merlin MWIR: Image saturation on a hot blacktop road

Fig. 6. Loss of stereo data will occur where there is low texture in a scene. The left example contains a frame from the 4m/s sequence in Fig. 5 where there was low texture on the dirt trail. Low texture regions can occur with cooled MWIR cameras as well. In the right example, the temperature extremes between the sky and blacktop road on a sunny day caused image saturation on the blacktop.

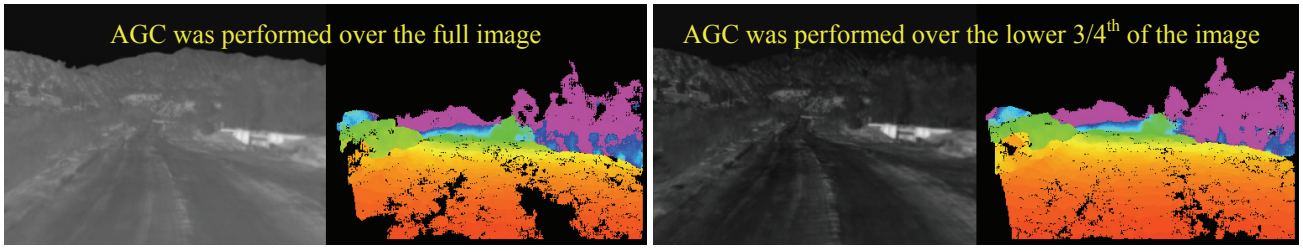


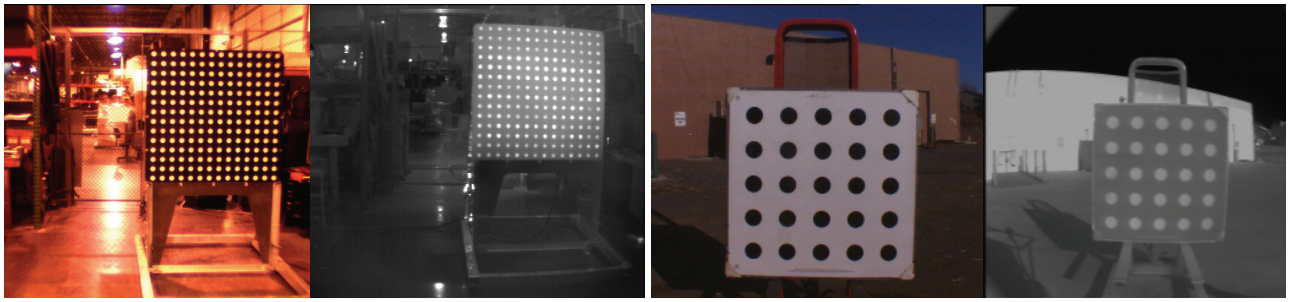
Fig. 7. A UGV was manually driven over a dirt trail lined with vegetation several times at 12 m/s with the AGC settings of a pair of Miricle 110KS cameras varied. Performing AGC over a portion of the scene that does not include the sky reduced the stereo loss due to low texture.

2.2 Calibrating a stereo pair of TIR cameras

It is necessary to establish calibration parameters for optically captured imagery, in order to remove distortion effects that are present as the result of camera design compromises. This process is termed, "Intrinsic Calibration". For stereo or multi sensor imagery, it is also necessary to establish the spatial relationship of the "two cameras", in order to provide the appropriate baseline for structural determination of range from stereopsis within the image pair. This process is termed, "Extrinsic Calibration". The most commonly used method for obtaining these calibrations is to collect images of a dot or checkerboard target, placed at various poses within the camera's FOV, and then to utilize techniques such as Bouquet's "Complete Camera Calibration Toolbox for Matlab®" to extract the desired transformation coefficients. For visible light cameras, capturing images of a calibration target is a simple matter. However, when one is calibrating TIR cameras, it is more difficult to obtain consistent calibration imagery. Additionally, if one desires to fuse visible and TIR imagery, it is desirable that both visible and TIR imagery be collected simultaneously with the two modalities of cameras viewing the same target in its various poses.

During Demo III, JPL calibrated stereo pairs of color and TIR cameras in a high bay with a target that contained an array of white circular plastic inserts. A miniature light bulb was taped to the back of each insert to heat them. During PerceptOR, JPL calibrated stereo pairs of color and TIR cameras with a target that contained black dots printed on a foam core board attached to an aluminum honeycomb panel. Positioning this target facing the sun for a few minutes was sufficient to generate enough thermal contrast between the dots and the background. There were issues with calibrating TIR cameras with both of these targets (Fig. 8). The plastic inserts in the lab target did not heat uniformly and the portable target was unusable on overcast days.

During RCTA, coauthor Gary W. Sherwin of GDRS developed a system that provides a robust multi-modality (visible, TIR, LADAR, RADAR) calibration target. As illustrated in Fig. 9, the target is a structure, supported by a 1.22 x 1.52 meter honeycomb aluminum panel with a thin insulating layer laminated to its front surface. A checkerboard pattern is printed on the back of a thin transparent lexan® sheet. This pattern provides the visible light stimulus of the target. Printing on the back of the sheet, allows the surface of the sheet to present a consistent material surface to retro-reflective sensors such as LADAR and RADAR. The TIR stimulus is provided by a grid of fine wires which is aligned with the transition lines on the back of the checkerboard. The wires run along the black, white transition lines and cross at the corner intersections of the visible pattern. When electrical current is passed through these wires, they heat up and shine TIR through the visible pattern and the covering sheet. The sheet is laminated atop the insulating layer of the target support. The resulting structure provides a precisely aligned multi sensor modality target which is relatively immune to external temperature effects, and can be used in a wide range of environmental conditions.



Lab: Black metal frame with white plastic inserts; a miniature light bulb is taped to the back of each insert

Portable: Black dots printed on foam core board attached to a rigid aluminum honeycomb panel; heated by the sun

Fig. 8. Previous targets used to calibrate both color and TIR cameras in the lab and field. The inserts in the lab target did not heat uniformly (left). The portable target worked for TIR cameras only on sunny days.

The calibration data collection procedure was a formalization of that described in the Matlab® Camera Calibration Toolbox. The toolbox obtains calibration information by locating the intersections of the corners of the checkerboard. The original tools require a great deal of manual processing to locate the bounds of the calibration target. GDRS developed additional algorithms to provide for fully automated operation of the calibration tools, and to locate both the visible checkerboard corners, and the TIR shining wire intersections. These tools make use of gradient information to find the location of the corners and intersections with high precision. Because of our use of gradients, it is advisable to avoid any condition that might bias the gradient maxima away from their proper location. Fig. 10 shows the prescribed

set of positions and poses used for obtaining calibration using the GDRS tools. Note that the various poses present a balanced set, with equal numbers of instances pointing left, right, down, up and orthogonally forward. This is done to cancel out any directional biases that might be present.

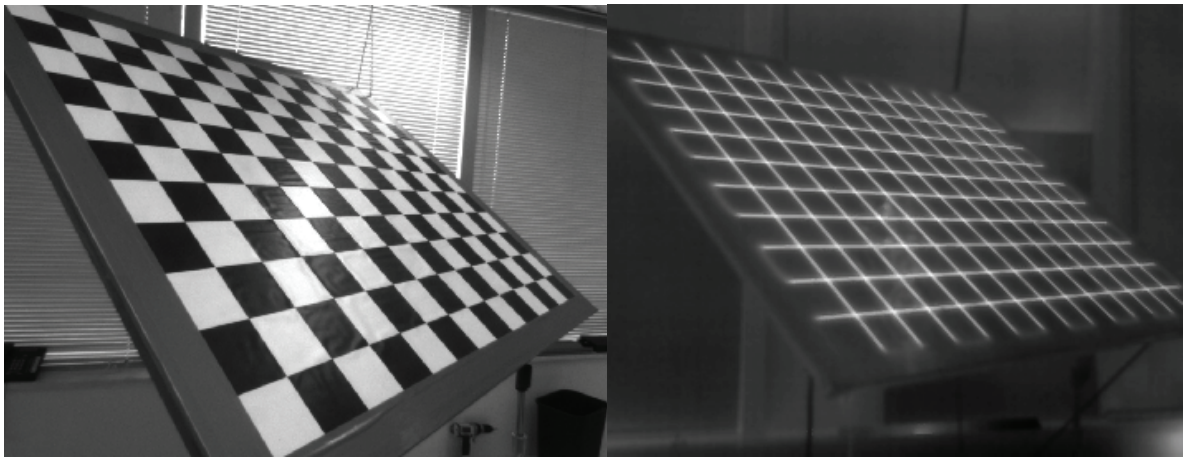


Fig. 9. A color (left) and Photon 640 LWIR (right) image of the GDRS multi-modality calibration target.

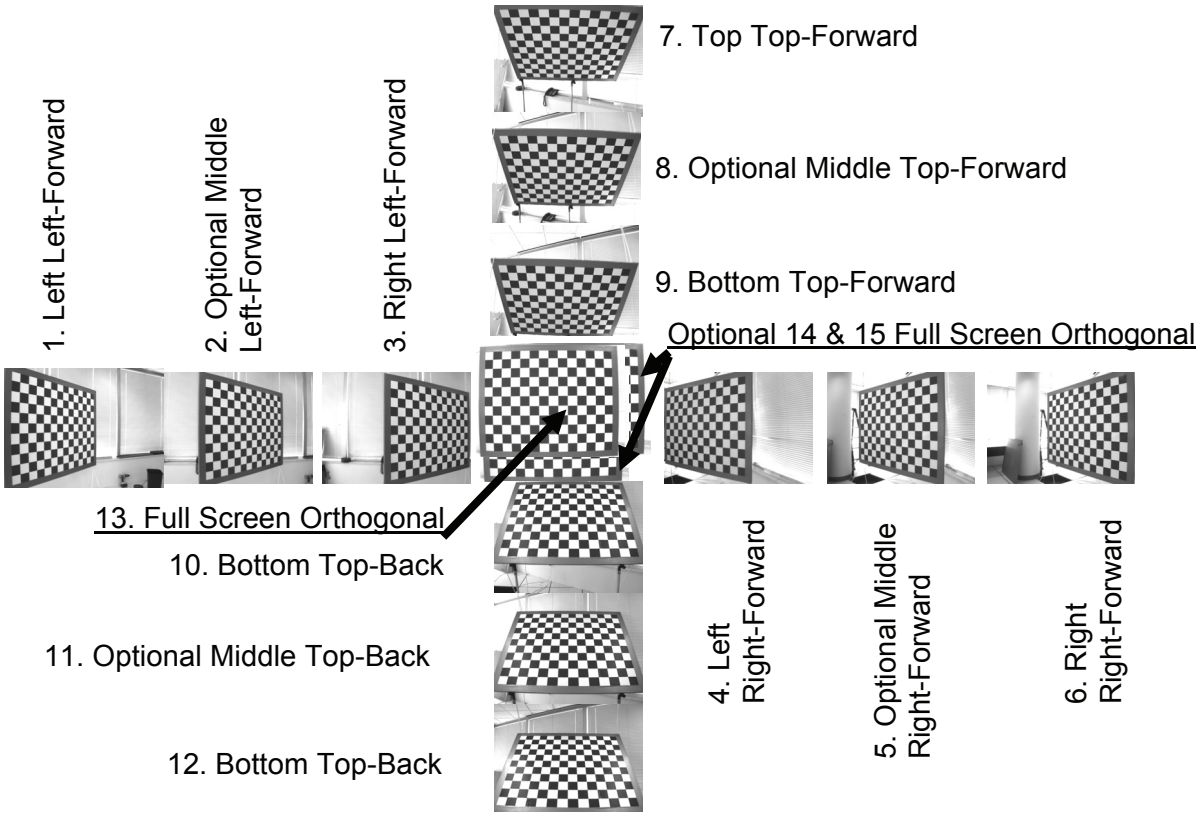


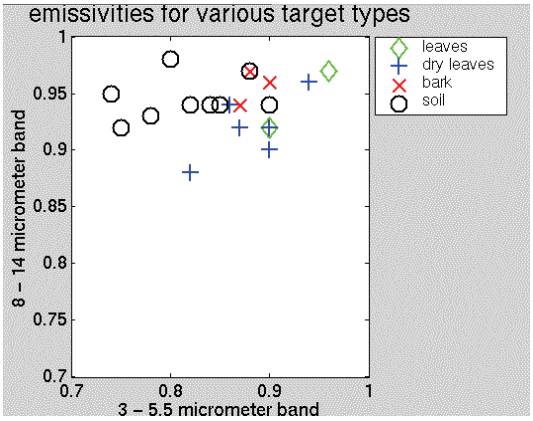
Fig. 10. Completion of the calibration process was via the routines called by the toolbox's calib_gui( ) and stereo_gui( ) associated routines from the Matlab® Camera Calibration Toolbox.

3. TERRAIN CLASSIFCATION

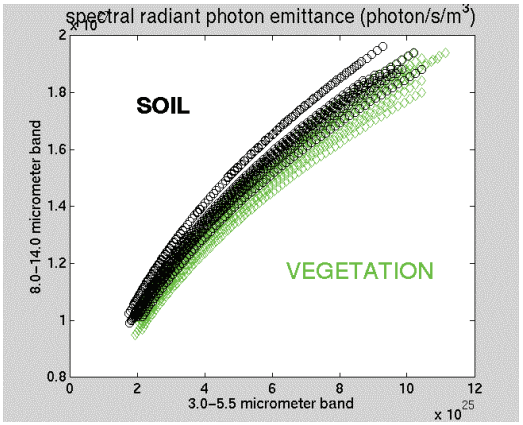
During PerceptOR, JPL conducted a number of experiments to determine whether real-time algorithms could be implemented for TIR-based terrain classification in outdoor environments. The primary application focused on day and

night traversals of terrain by autonomous UGVs. One of the motivations for terrain classification comes from the realization that range sensors alone may not adequately distinguish between non-traversable obstacles (rocks) and traversable obstacles (small bushes.) We concluded that daytime terrain classification with visible and near infrared imagery was becoming reasonably mature while classification at night remained a problem.

We considered emissivity signatures as a potential way to distinguish vegetation from soil, rock, and other materials. Temperature, in part, is a measure of emitted radiation as a function of emissivity and that presented a problem. Each material has an unknown emissivity for each spectral band plus an unknown temperature. (For *N* bands of multispectral data, there are *N+1* unknowns; temperature plus *N* emissivity). While normalization algorithms for estimating emissivity from airborne and orbiting instruments exist for well calibrated sensors containing 5 bands or more. No such instruments, however, were practical for UGV applications. We adopted the following strategy instead: As temperature varies, each material will cluster in different regions of infrared color space making it is possible to distinguish a few terrain classes without explicitly estimating material temperature or emissivity. In particular, vegetation and soil/rock have sufficiently different emissivities in the broad MWIR and LWIR bands (Fig. 11).



(a) MWIR and LWIR emissivities for soil and vegetation materials.



(b) MWIR vs. LWIR radiance as a function of temperature (0 to 40ºC) for the emissivities in (a).

Fig. 11. Strategy for TIR terrain classification.

3.1 Single Band TIR Terrain Classification

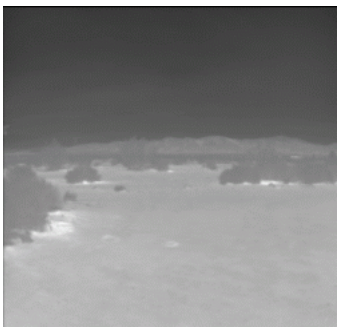
The thermal emission characteristics of certain terrains thus exhibit behaviors that can be exploited to perform automatic terrain classification. The ability to discriminate soil from vegetation provides significant advantages for autonomous navigation. Small to moderate size obstacles that can be verified to correspond to vegetation enables more agile traverse of the terrain, by allowing the UGV to avoid having to drive around obstacles that can be run over without damage. The observed thermal emission behavior in arid and semi-arid terrain is one such case. Vegetation appears consistently darker than the soil during the day. At night, the emission behavior reverses. The periods of relative stability are diminished only during sunrise and sunset, where the contrast between vegetation and soil reverses.

To characterize the ability of a single band terrain classifier for this type of terrain, we performed a data collection in the Calico desert near Yermo, California. A primary goal was to differentiate vegetation from soil. Our earlier work on terrain classification from a single MWIR band from similar terrain showed very good results during the night. In this experiment we focused on daytime classification, exploiting the characteristics of thermal emissions from vegetation and soil in this type of terrain. Namely, we can expect that soil emits a higher level of MW IR radiation during the day than the vegetation does. The behavior reverses during the night, with two crossovers, one at sunrise, and the other at sunset. The relatively stable environmental conditions in arid and semi-arid terrain are expected to affect emissions in a reduced manner, and thus, make these types of scenes amenable to the application of simple classifiers that can be implemented to run in real-time.

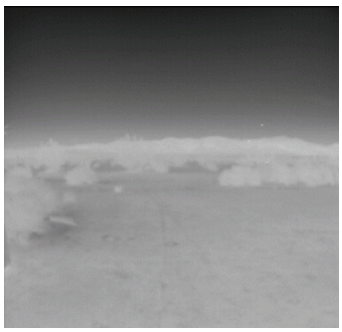
Fig. 12 illustrates the chosen scene. We acquired 93 stereo pairs of color and NC256 MWIR images, one set every 15 minutes, starting at 3:00 pm on March 25, 2003. The cameras were positioned facing north, towards the Calico Mountains. Two small ant hills can be seen to the lower left in some of the MWIR imagery. A portion of a concrete

parking space head stone can also be seen to the lower left of the MWIR imagery. There are two types of vegetation in the scene: a large bush approximately 2.3 meters tall, 2.3 meters wide, not very thick, with small green leaves; a shorter bush, dense, very dry, approximately 0.76 meters tall. The distance to the bushes is approximately 25 meters from the camera position. There are two small stumps about 16 meters from camera position, one at 11 o'clock and one at 12 o'clock. They are actually part of a tree-line that was cut down that runs East-West. The soil in the foreground has virtually no grass, while soil in the background has a very short grass carpet layer. There is a significant number of broken glass fragments in foreground. Sunset was recorded at 5:55pm and sunrise at 5:45am. First light appeared on the eastern sky at 4:55am. There was a quarter moon in the South-Eastern sky. The weather report at 10:25am indicated winds from the West at 28 mph, gusting up to 36 mph; at 12:08pm winds from the West at 33 mph, gusting up to 43 mph, blowing sand on sensors and equipment (Fig. 13).

(a) Calico desert scene, near Yermo, California.



(b) MWIR Daytime frame 48 (12:00 pm)



(c) MWIR Nighttime frame 80 (8:45 am)

Fig. 12. Calico desert scene and representative left MWIR camera frames.

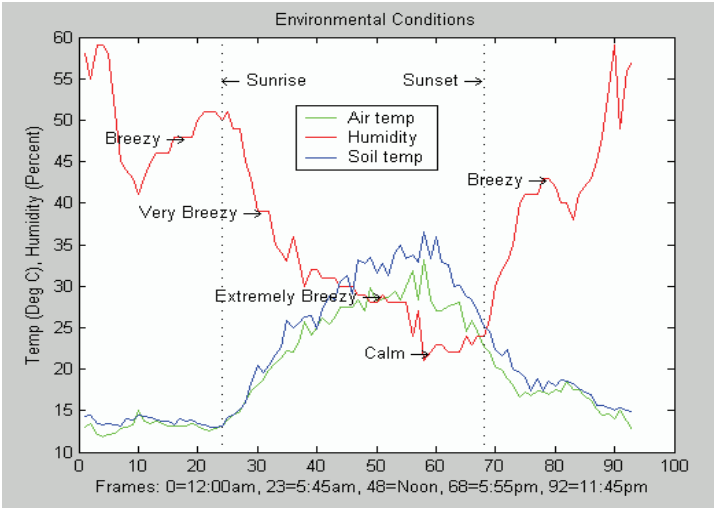


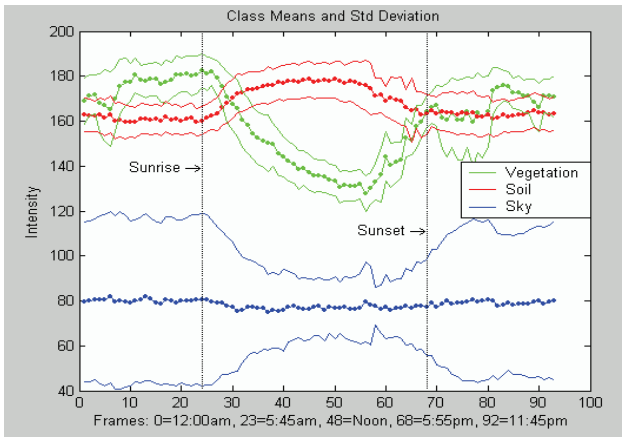
Fig. 13. Environmental conditions prevailing during the Calico desert data collection.

A primary concern in developing perception software for UGVs is the reasonable expectation that they will run at near real-time speed. Simplified algorithms lack the sophistication of more elaborate methods available and, trading accuracy for speed. We tested two algorithms for single band classification. The first method uses a training set to collect statistics. The training statistics vary from simple first order statistics to weighted probability distributions. Each pixel is classified independently according to a distance measure. The supervised classifier uses means and standard deviations to calculate the Mahalanobis distance of each pixel to the training set. Our second method was unsupervised. It first segments the image in a predefined number of clusters. This process requires a number of iterations to update the clusters around the means. The final classification for these regions is given by nearest distance to reference mean values. These relatively simple single band classifiers can be applied to the imagery under reasonable stable environmental periods. We have also reported much more complex thermal behaviors (see 3.2). The influence on

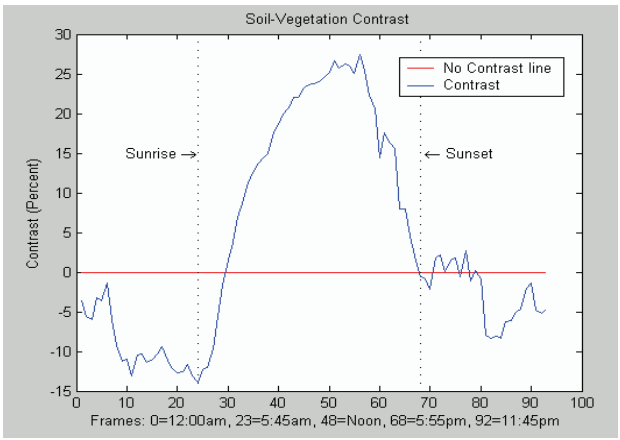
recorded thermal emissions from complex weather conditions, and complex scenery is not well understood yet, and requires further study. Also, the collected detailed environmental data was not used explicitly to help explain thermal behaviors. Instead, we chose the daytime frames to test the two classifiers described.

Although we can expect that environmental conditions have an influence on surface temperatures from convected heat, our sensors acquire measurements of emissivity, i.e. radiated heat. Intuitively, medium density (from water or particulate content) may have only minor effects of photon densities in scenes like these. Multiple reflections from objects in the scene, clouds in the sky, and large bodies of water will add up in complex ways. These can be characterized accurately only when the mapped intensities reflect photon count and have not been altered by a sensor's AGC. AGC is useful for viewing and for stereo matching, however, and was applied to the imagery in the Calico desert dataset.

Fig. 14(a) shows the means and standard deviations of regions corresponding to vegetation, soil and sky. The template, or mask, used to collect these statistics was easily constructed manually using frame 48 shown earlier in Fig. 12(b). Fig. 14(b) illustrates the relative contrast between soil and vegetation. These curves help visualize the precise time of the expected thermal crossover for these two classes, and determine the extent of the transition periods with respect to the sunrise and sunset times. In the experiments below we processed the daytime frames from 32 to 62, just inside the thermal crossovers near sunrise and sunset. The daytime mean for 3 million plus samples of soil is 176.2, and for almost a million samples of vegetation is 139.6. These fall well outside each other's standard deviations to provide training values for very good supervised or unsupervised classification.



(a) Statistics for three classes: soil, vegetation and sky.



(b) The available contrast between vegetation and soil.

Fig. 14. Statistics for training the classifiers collected over the heating cycle. Both supervised and unsupervised classifiers use the statistics in (a). The unsupervised classifier does require reasonable contrast between the material classes.

Fig. 15 illustrates results for frame 48 (noon). The thermal image is shown in Fig. 15(a). The bright areas under the bushes are protected from the wind and thus remain warmer than the surrounding soil. The supervised classifier was trained with all the daytime samples and produces the result shown in Fig. 15(b). The unsupervised classifier used seven iterations to converge to the six segmented regions shown in Fig. 15(c). Using the same means computed for supervised classification, the obtained unsupervised classification result is shown in Fig. 15(d). Unsupervised classifier results for scenes earlier and later than frame 48 (noon) are illustrated in Fig. 16.

3.2 Dual Band TIR Terrain Classification

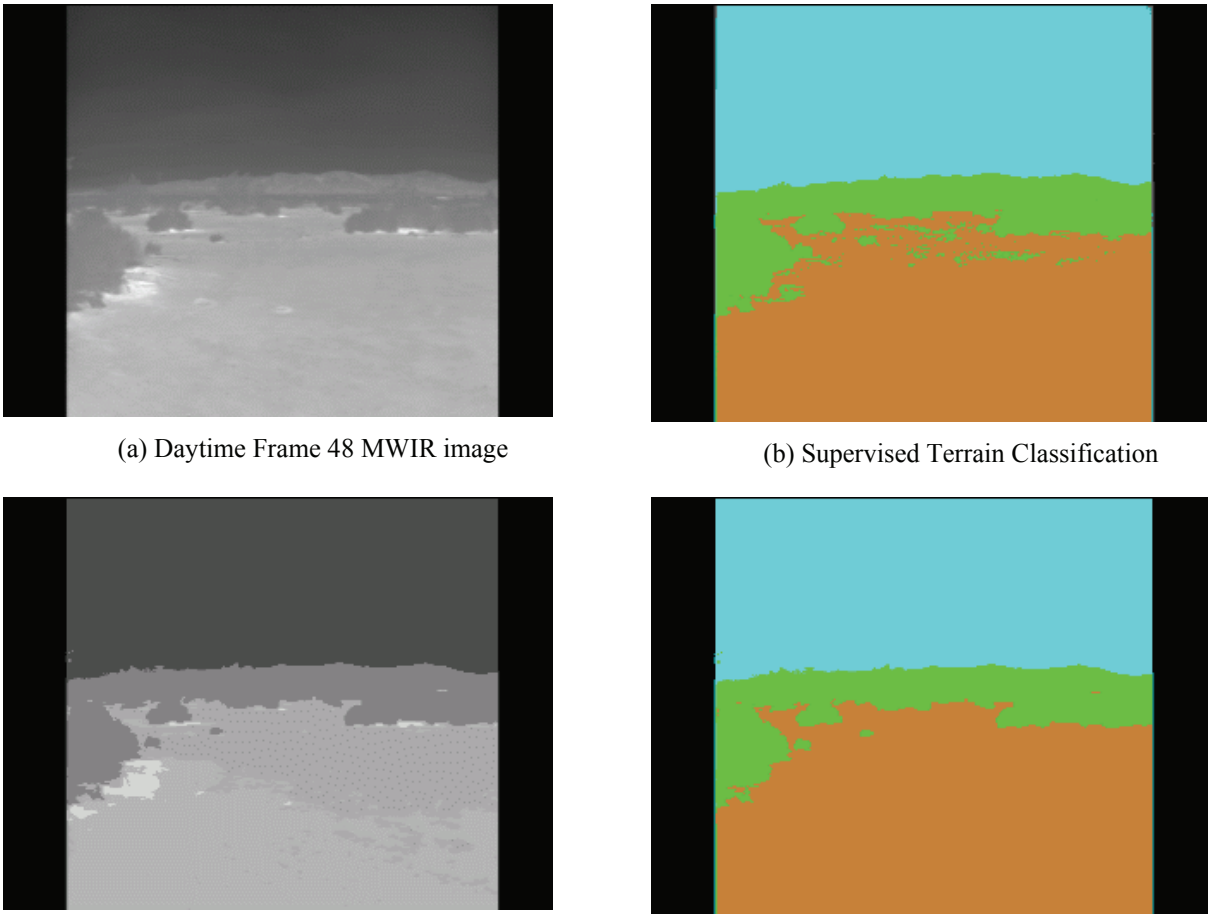
During PerceptOR, the passive perception system for the team Raptor UGV included a stereo pair of color and NC256 MWIR cameras and one Alpha LWIR camera. We registered the MWIR and LWIR bands by backprojecting the range data from the MWIR stereo pair onto the LWIR image. When range data was not available, we had registered bands by using affine warping with manually designated corner points. Fig. 17(a) illustrates the stereo rig and the scene in the Arizona desert where the data collection was carried out, Fig. 17(b) shows an image acquired by the LWIR camera at 9:30pm, and Fig. 17(c) and Fig. 17(d) show images from the MWIR camera at 3:30pm and 9:30pm, respectively.

We used two automated approaches for generating nominal ground truth classification. First, since vegetation is distinctly brighter than the ground at night, we generated image segments by manually choosing a threshold from a single band image. Secondly, using time-lapse static-scene imagery, we computed and thresholded the brightness ratio of an early and late image. Both methods produced comparable results.

The ground truth classification images in Fig. 18 were used as masks to index and label the 2-D histograms of the MWIR/LWIR imagery shown in Fig. 19. The histogram in Fig. 19(a) is representative of the histograms for all images taken after sunset and shows good performance as the class clusters are reasonably separated. To register the two bands in this experiment we used an affine warp which yields approximate registration. The observed cluster mixing could be an artifact of registration error in the affine warp. The daytime histogram in Fig. 19(b) shows progressively more mixing of the clusters the earlier in the day the images were taken. For the most part, the "peaks" in this color space remain well separated. Fig. 19(c) illustrates the peaks in the histograms for all the imagery taken between 2:30pm and 9:30 pm.

The Mahalanobis distance measures how close two populations are statistically. The classification results for the 9:30pm MWIR/LWIR imagery using a Mahalanobis distance classifier trained on all of the post-sunset imagery is shown in Fig. 20. Note the comparison to the ground truth segmentation shown above in Fig. 18.

Another experiment was conducted under more complex environmental conditions using LWIR and MWIR imagery collected at Chatfield State Park, Littleton, Colorado. JPL/SAIC collected a 24-hour time-lapsed (every 20 minutes) movie from a stationary position of a scene containing soil, vegetation, water and clouds. Fig. 21 shows the scene and the starting left MWIR camera and LWIR camera frames at 8:00 pm. The cameras had different focal length lenses and the correspondence between the pixels was performed by affine warping with manually designated corner points.



(a) Daytime Frame 48 MWIR image

(b) Supervised Terrain Classification

(c) Six-cluster K-means Segmentation

(d) Unsupervised Terrain Classification

Fig. 15. Single band TIR terrain classification result for frame 48 (12:00 noon).

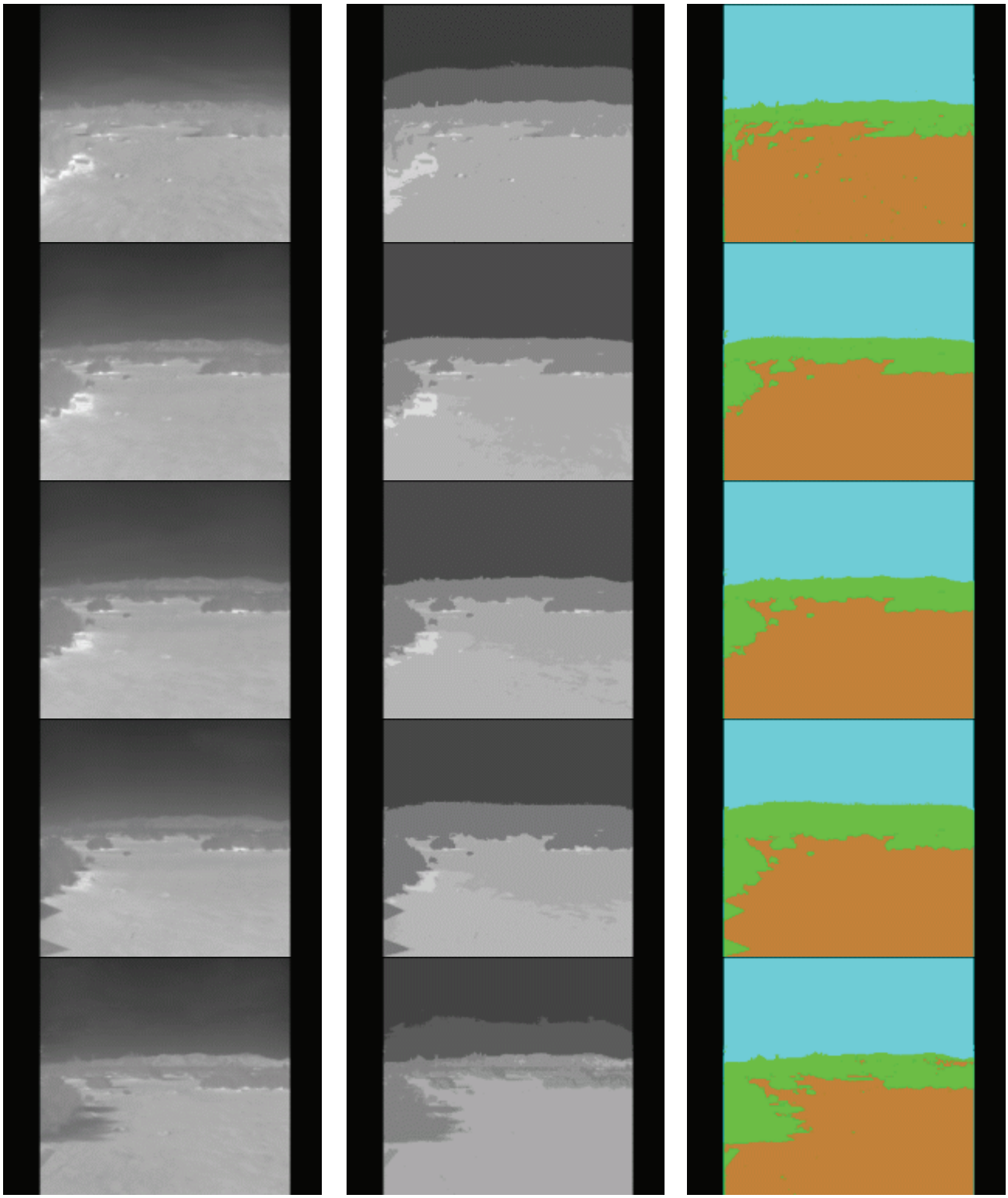
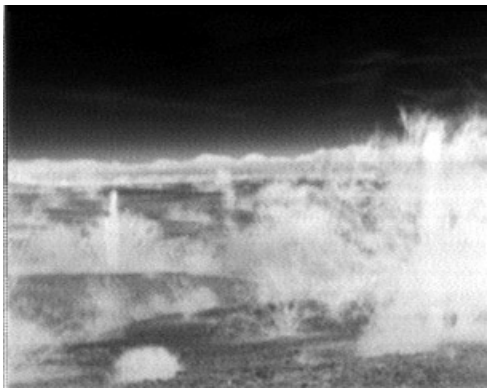
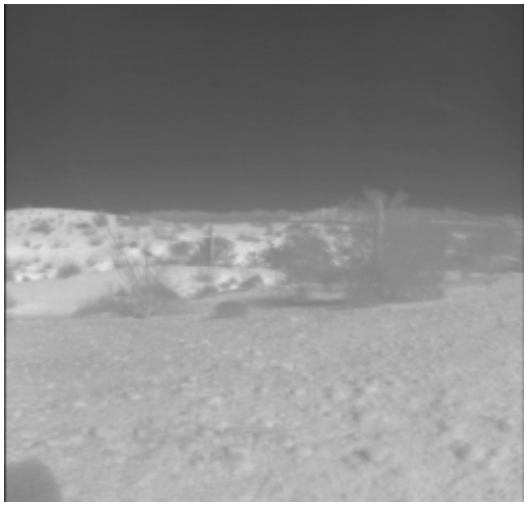


Fig. 16. Unsupervised terrain classification results for daytime frames 32, 40, 48, 54 and 62. The left column shows the TIR image, the middle column shows the 6-means automatic segmentation, and the right column shows the classification results for soil, vegetation and other (sky.)

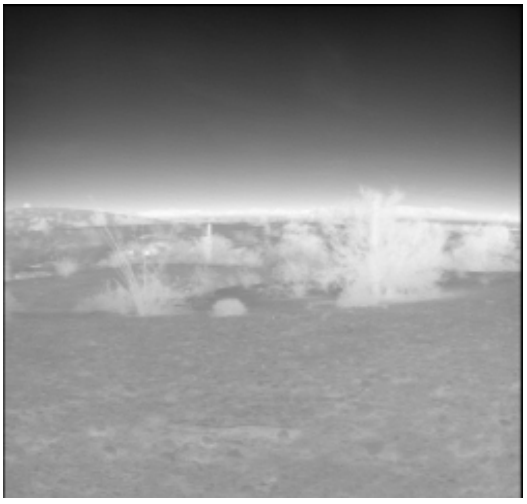
(a) Color image of scene and sensor setup. Two NC256 MWIR and one LWIR cameras were used.



(b) Unwarped LWIR (10.5-11.2μm) filter wheel camera image at 9:30pm. The image contrast is stretched for illustration.

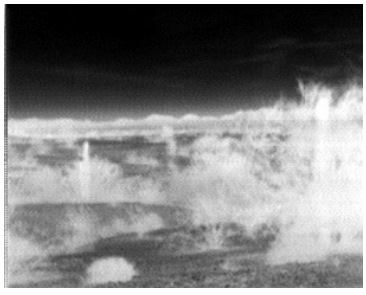


(c) NC256 MWIR image at 2:30pm with AGC on.

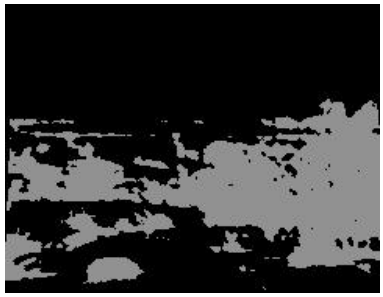


(d) NC256 MWIR image at 9:30pm with AGC on.

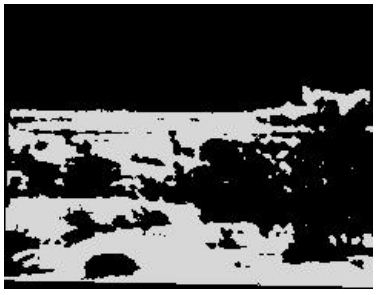
Fig. 17. Data collection at the Arizona desert for dual band TIR terrain classification. The vegetation is quite dry. The ground is sandy and stony. The vegetation tends to be cooler (hence darker) than soil during the day and the reverse at night.



(a) Unwarped LWIR filter wheel camera image at 9:30pm.

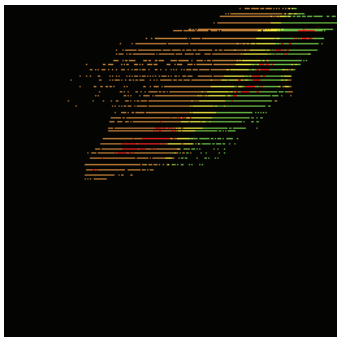


(b) Bright areas are ground truth vegetation regions.

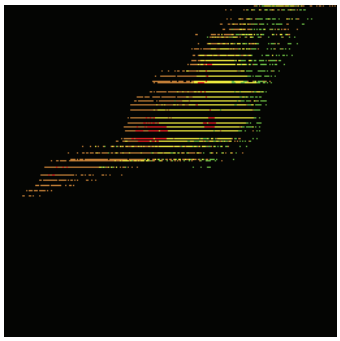


(c) Bright areas are ground truth soil regions.

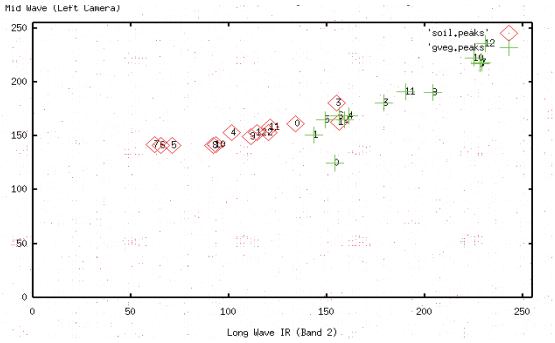
Fig. 18. Ground truth for evaluation of classification performance.



(a) 2-D histogram of soil and vegetation regions for the 8:30pm imagery.



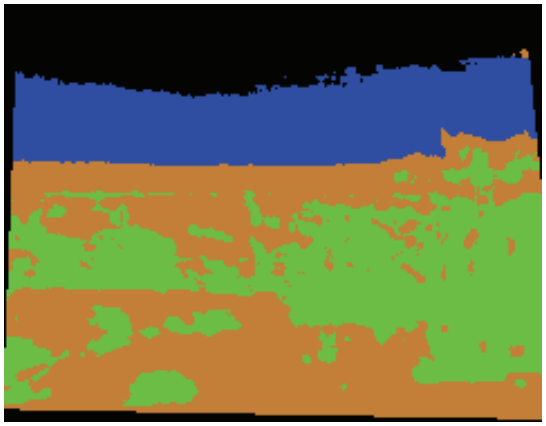
(b) 2-D histogram of soil and vegetation regions for the 4:30pm imagery.



(c) Plot of the peaks of the 2-D histograms for the soil and vegetation clusters for all the imagery taken between 2:30pm to 9:30pm.

Fig. 19. Two-D histograms show clustering of soil (brown) and vegetation (green). The horizontal axis represents LWIR and the vertical axis MWIR. Cluster overlap or mixing is shown in yellow. Cluster areas in red show the "regional peaks", or, areas for each cluster with largest pixel counts. In (c) most peaks are well separated in color space. Labels indicate peak location in the sequence with lower labels representing earlier times and higher temperatures.

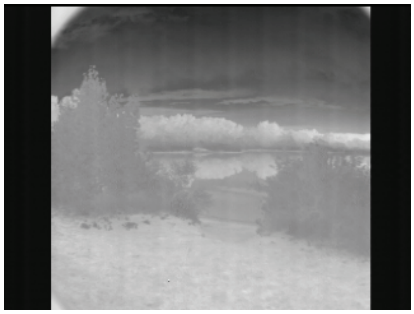
(a) Dual band classification at 4:30 pm.



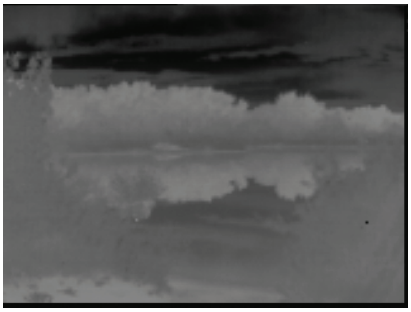
(b) Dual band classification at 9:30 pm.

Fig. 20. Classification results for the 4:30pm and 8:30pm MWIR/LWIR imagery using a Mahalanobis distance classifier. For dayttime training we used the two earliest pairs. For nighttime we used the eigth latest pairs for training.

(a) Color image of scene at 4 pm.



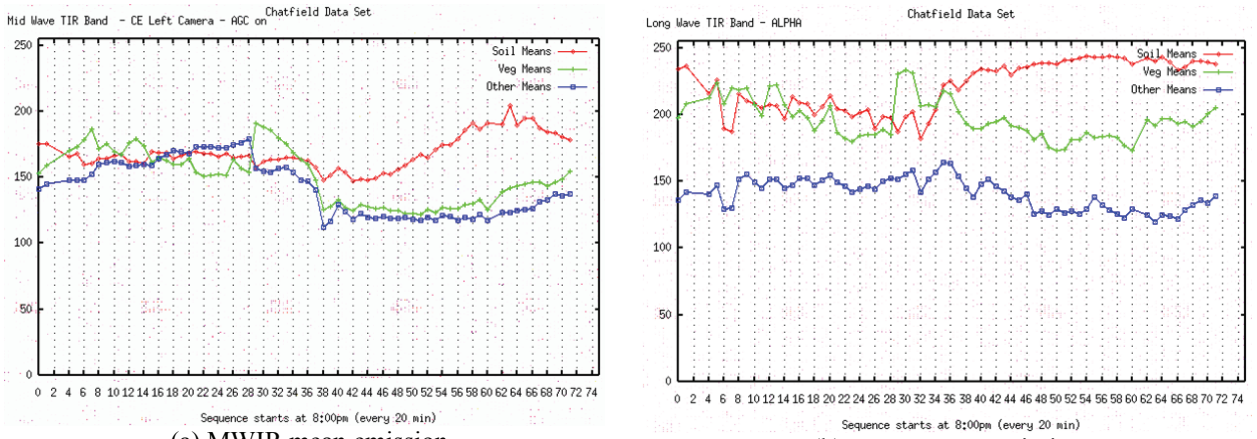
(b) MWIR image at 8:00pm



(c) LWIR image at 8:00 pm

Fig. 21. Sample nighttime (8:00pm) MWIR and LWIR images of the data collection scene.

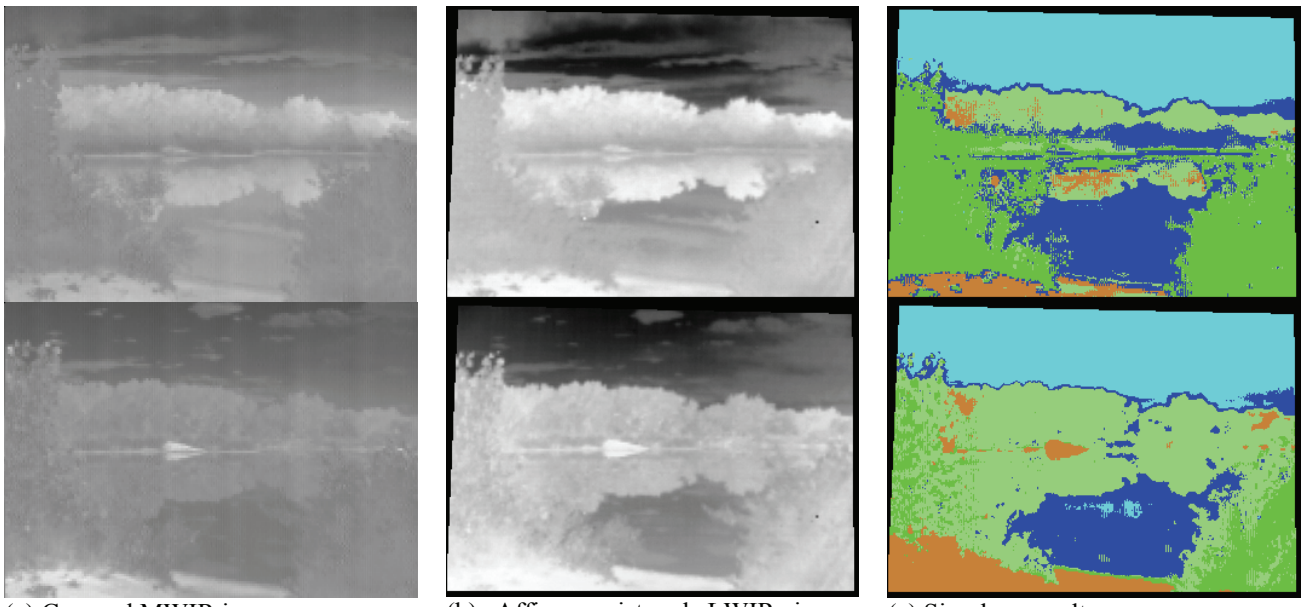
The statistical information for soil and vegetation for the entire sequence is shown in Fig. 22. The statistical means of radiated heat for soil and vegetation over a 24 hours thermal cycle show that only during the daytime vegetation and soil emissions appear consistently distinguished. Note that in the experiment above with data collected in the Arizona desert, the clusters in the 2-D histograms for images acquired after sunset separate fairly well in color space. The nighttime classification results were very good when qualitatively compared to ground truth masks created by manual segmentation. The mixing occurred during the daytime. In this experiment, Fig. 22 shows the mixing occurring during the nighttime, an indication of the influence of the type of terrain and the different environmental conditions. For both experiments, the camera's AGC was on to favor stereo processing. When automatic exposure, gain, or level controls are used, they will shift the clusters and the emission means by unknown amounts. This introduces a new problem of how to track the cluster and means changes over time. Fig. 23 shows classification results for frame 0 (8:00 pm) and frame 65 (5:40 pm) using six classes to attempt to differentiate different types of vegetation and water and sky as well.



(a) MWIR mean emission

(b) LWIR mean emission

Fig. 22. Statistical means of radiated heat for soil and vegetation over a 24 hours thermal cycle. Only during the daytime vegetation and soil emissions appear consistently distinguished.



(a) Cropped MWIR image

(b) Affine registered LWIR image (stretched for illustration)

(c) Six-class result.

Fig. 23. Terrain classification of frame 0 at 8:00pm (top row) and frame 65 at 5:40pm (bottom row). Note that the vegetation reflection on the water surface is also classified as vegetation as the water acts like a mirror.

A number of recommendations emerged from the terrain classification experiments: A major source of error is the approximated registration of the two TIR bands. Cameras are now available with MWIR and LWIR detectors at each pixel that would provide the best registration, and their use is the recommended follow on approach. A second aspect of improvement is in modeling image noise to determine how much spread noise causes in the color space clusters. Also, the need for exposure, gain, or level control must be evaluated. If these are needed, strategies must be developed to track the effects they have on class statistics over time. In addition, additional datasets are needed, acquired under wider ranges of conditions, including different meteorological conditions and different terrain. Finally, experiments with high resolution laboratory emissivity spectra have suggested that more classes could be discriminated if more bands are available. This suggests more in-depth studies to determine what bands produce the best separation, how to process the data to do terrain classification, and how to design and build practical sensors for UGV-mounted applications.

4. TREE TRUNK DETECTION

During PerceptOR, JPL developed real-time algorithms to identify tree trunks in the view of a UGV [6]. The work was designed to estimate the diameters of trees that belong to a large class of trees that have nearly straight, nearly vertical trunk support. The underlying assumption for the detection of the tree trunks was that sections of their main trunk, however visually fragmented, are discernable from the background. These portions are required to have detectable corresponding boundaries that delineate portions of the trunk. We called these portions, trunk *fragments*. Our main interest was in accurately characterizing trees that are close to the UGV, e.g., within 15 meters. In particular, the portions of trees that are within the height of a UGV are crucial to detect.

Different types of trees will have different appearances depending on the texture of the bark, the smoothness of the exposed core, and the density of the canopy. Moderate to dense canopies will reduce the amount of direct sun illumination. Direct illumination from certain angles in combination with the viewing angles can result in illumination discontinuities that can make the accurate detection of reliable tree fragments difficult. During the night however, TIR emissions are very stable, and in the absence of rain, most measurements were, and should, be very accurate.

Our real-time algorithm implementation made no special effort in combining detected fragments into complete tree trunks but does extract as many fragments as possible, even if they correspond to trees in the distance, or trees that may be occluded by other trees or other objects. Perceptual curvilinearity, however, can be applied if more complete tree trunk models are needed. The details of the algorithm are given in [6] so we only summarize them here and provide a discussion and examples from TIR image analysis. Note that we used stereo pairs to obtain cues from range data to estimate the diameters of tree trunks from their 2D image projections. Typically, the range data is provided relative to the left image in the stereo pair, and the detections are from this image.

4.1 Tree Trunk Algorithm

We tested the algorithm on NC256 MWIR stereo pairs acquired at Fort Polk during the night. The phenomenology indicates that tree trunks appear brighter than their surround as they have thermal inertia properties that radiate heat at lower rates. The algorithm consists of the following steps:

4.1.1 Edge Detection and Contour Extraction: Intensity edge detection is carried out using the Canny edge detector, which also incorporates edge-thinning and contour feature extraction. Fig. 24(a) illustrates an example. Note that some of the edge boundaries of portions of the upper half of the trees are difficult to detect due to saturation problems during image acquisition but sufficient information is available to form contours and extract tree fragments. The same problem affects the density of the range map. This example, however, shows that the method tolerates partial information. Note that the thinnest trees in the distance may be difficult to discern as their 2D widths are at, or below, the ability of the edge detector to resolve. The edges extracted from the image are illustrated in Fig. 24(b). The thin stock from the tall grass in the foreground, however, are close, and their boundaries can be resolved well. The directed, linked edge contours extracted are shown in Fig. 24(c).

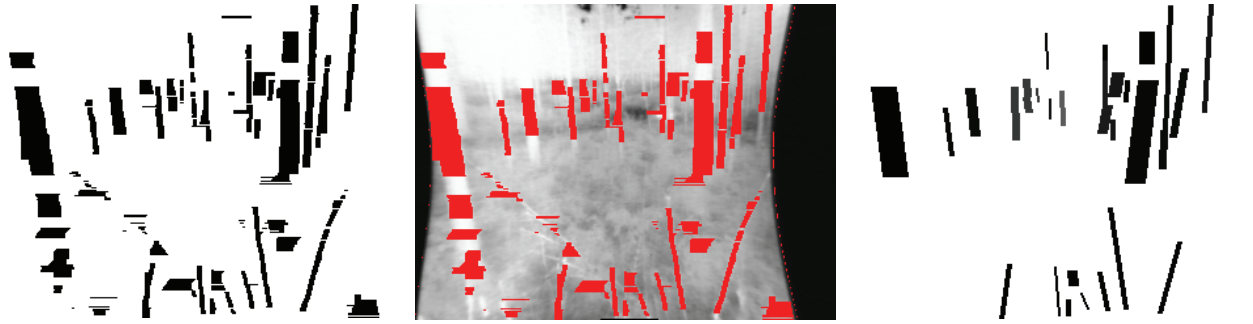
4.1.2 Tree Trunk Fragment Extraction: Matching the corresponding edges along these directed contours uses the thermal cue that trees appear bright at night. In this example, regions having intensity higher than the mean image intensity help indicate the presence of trees and other kinds of vegetation. The resulting raw fragments are shown in Fig. 25(a), and also overlaid on the input image for illustration in Fig. 25(b). The selected fragments represented by linear segments are shown in Fig. 25(c). Note that the thin tall grass stock in the foreground is well represented.

(a) NC256 MWIR rectified image (left camera)

(b) Thinned filtered edges from image at left

(c) Edge contours color denote edge direction (bright side)

Fig. 24. Sample edge detection and contour extraction result.



(a) Raw fragments mask

(b) Raw fragments in (a) overlain on image in Fig. 24

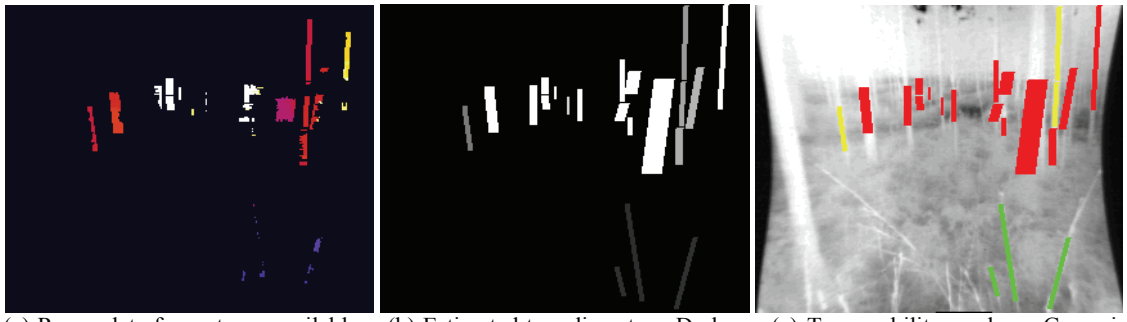
(c) Linear segments approximating selected fragments

Fig. 25. Tree trunk fragment extraction by matching corresponding edge contours and representing connected components by linear segments.

4.1.2 Diameter Estimation: In order to estimate the diameter, D, of the tree fragments we use the range image computed by the stereo system. By similar triangles:

$$D = \frac{d\ R}{f} \tag{2}$$

where d is the 2-D projection of the diameter on the image, f is the focal length of the camera lens, and R is the distance, or range, to the tree. Fig. 26(a) illustrates the range data available for the tree fragments. We required that at least 5 percent of the area of the fragment have range information available. To determine the range to the tree (R in the equation above) we compute the average range in the center 2/3 portion of the area inside of the linear tree segments. This is to minimize range data contributions near the boundaries of the trees, where the range values may incorporate background range values. The average ranges are illustrated in Fig. 26(b). A traversability assessment can now be made as a function of the vehicle capabilities. Fig. 26(c) illustrates the tree trunk models overlain on the image, where green fragments denote the vehicle can traverse over the object while red means it cannot. The large tree on the left, although detected, does not have range data associated because it falls outside the right image of the stereo pair but it is safe to assume that it was seen earlier when the vehicle was at a slightly larger distance from it. An analysis on the accuracy of the diameter estimates compared to ground truth measurements is given in [5].



(a) Range data from stereo available for distance to raw three fragments. Blue is closer, white is farther.

(b) Estimated tree diameters. Darker is thinner.

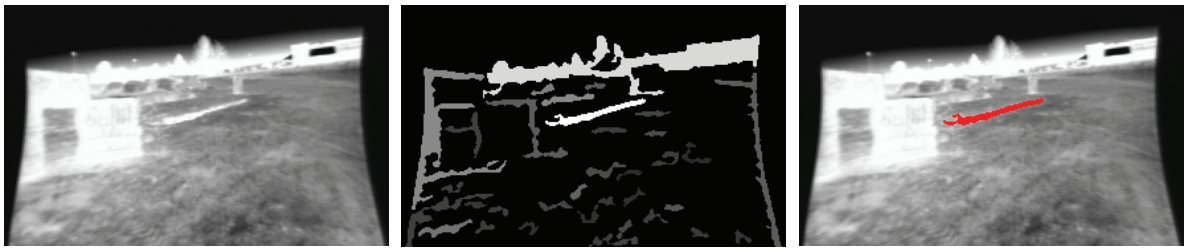
(c) Traversability markers. Green is traversable by vehicle, red is not.

Fig. 26. Tree diameter estimation and traversability assessment for an all-terrain vehicle sized UGV. In the right image, we predict the trees with green and yellow markers can be pushed over by the UGV.

5. NEGATIVE OBSTACLE DETECTION

Detecting negative obstacles that are narrow (yet wide enough to cause UGV damage) is a particularly challenging problem. One reason for this is the visible portion of a narrow negative obstacle shrinks rapidly with increasing range, causing them to be difficult to detect in time to avoid at cross country speeds that most UGV programs aspire to achieve. During PerceptOR, JPL investigated if a thermal signature exists that could be exploited to increase the range of detection of narrow negative obstacles. A 24-hour data collection was performed on a 0.53 meter wide L-shaped trench in soil. NC256 MWIR stereo images were collected every hour from 6 different distances (ranging from 4.6 to 18.3 meters) along the four cardinal directions. For all four viewing directions, the interior of the trench remained warmer than the surround throughout the night (from 9pm to 6am) [7].

JPL has developed a negative obstacle detector that combines thermal and geometric cues [8]. A Gaussian operator and Laplacian operator are applied to the rectified left camera intensity image in a single step (LoG convolution). Gaussian smoothing effectively suppresses the influence of the pixels that are up to a distance sigma from the current pixel. The Laplacian operator approximates the second derivative and is non-directional. The output of the LoG convolution is used to produce a closed contour image. Stereo ranging is performed and each closed contour region is analyzed for thermal and geometric cues. The normalized mean intensity difference between the interior and exterior of closed contours are thresholded and the stereo range image is used to exclude "warm" objects that are not a part of the ground plane. Fig. 27 contains a sample result. At the maximum detection range (16.8 meters), the trench only spanned 3 rows in the intensity image. Typically, there needs to be at least 7 pixels (the height of the stereo correlation window) across the width of a negative obstacle to be geometrically detected in stereo range data. Further gains in detection range could be realized by using a TIR sensor with a smaller vertical angular pixel resolution. Uncooled TIR cameras are now available having one fourth the vertical angular pixel resolution (1.12mrad, 33° vertical FOV) as the cooled NC256 MWIR sensors used for this study.



(a) NC256 MWIR rectified image of a trench near SAIC, Littleton, CO

(b) Segmented, normalized intensity difference image for the scene in (a)

(c) Negative obstacle detection for the scene in (a)

Fig. 27. Sample negative obstacle detection result based on combining thermal cues from a rectified intensity image and geometric cues from stereo range data. In this sequence approaching a 0.53 meter wide trench from a range of 16.8 meters, the trench was detected in all 20 frames with no false positive detections.

6. WATER AND MUD DETECTION

Under RCTA, JPL has developed several daytime water detection algorithms that use as stereo pair of color cameras [10][11][12]. Nighttime water detection with a single sensor type is a more challenging problem. In the previous section, we saw that narrow negative obstacles tend to remain warmer than their surround throughout the night. We have also investigated if there is a thermal signature that can be exploited to detect water bodies during the night. During PerceptOR, JPL performed a 24-hour data collection from a stationary position at Chatfield State Park on a scene that contained soil, vegetation, a large water body, and a portion of the sky. A plot of the mean NC256 MWIR intensity values for the terrain types in the scene are shown in Fig. 28. The water body was warmer than the other terrain types from 11pm to 4am. The reason for this is the water body cooled at a much slower rate than the other terrain types since water has a high specific heat capacity. During most of the daytime (9am to 8pm), the soil was the warmest terrain type and the water body had a similar temperature to the vegetation. For this data set, there appears to only be a small portion of the night where a single band TIR sensor would be useful in detecting water bodies based on thermal signature.

But there are other ways to detect water bodies. We saw in Fig. 21 that specular reflections occur in both color and TIR imagery. As illustrated in Fig. 29, three cues for water bodies in color imagery are also manifest in TIR imagery; 1) specular reflections of objects in the background (such as trees) occur in water bodies [12], 2) water bodies tend to have a similar intensity to the sky at far range [10], and 3) water body intensity values decrease with increasing incidence angle at a rate higher than other terrain types at close range [11]. Therefore, all three of JPL's specialized water detectors are applicable to TIR imagery.

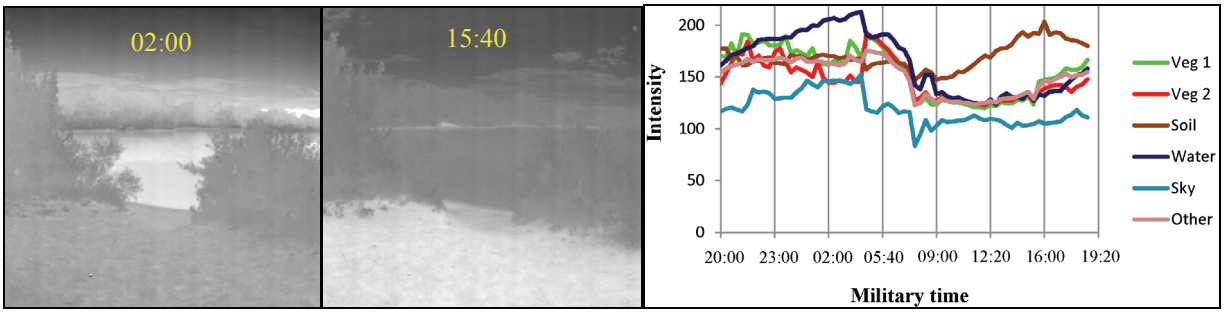


Fig. 28. NC256 MWIR statistical means of radiated heat for several terrain types during a 24-hour data collection from a stationary sensor position. In the graph, "Veg 1" and "Veg 2" curves correspond to the vegetation on the far and near side of the water body, respectively. The water body was warmer than the other terrain types in the scene from 11pm to 4am.

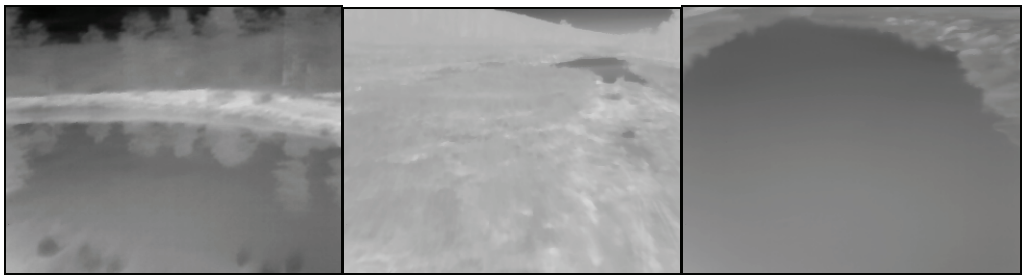


Fig. 29. Sample images from Photon 320 LWIR stereo data sets collected at Ft. Indiantown Gap in May 2008. Similar to color imagery, water bodies in TIR imagery reflect trees in the background (left), have a similar intensity to the sky at long range (middle), and have an intensity that decreases with increasing angle of incidence (right). Thus, all three of JPL's specialized water detectors are applicable to TIR imagery.

Specular reflections of objects in the background in still water can be detected during the daytime and nighttime in TIR intensity imagery by performing cross correlation between targeted pixels below and above the horizon, or within a stereo range image by locating pixels below the horizon that have a similar range as an object in the background [12]. An example of detecting a water body in stereo range data during the daytime (at 3:00pm) is illustrated in Fig. 30. In the stereo range image, notice the range to the portion of the puddle reflecting trees in the background is similar to the range to the trees in the background. In the traversability map, the puddle was first detected at a range of 12.0 meters during a

50 meter southbound approach using the JPL developed stereo reflection detector described in [12]. The Miricle 110KS LWIR stereo images in this sequence were collected with AGC being performed over the entire image in peak mode. Note that there are a lot of low texture regions on the dirt trail and on the tall vegetation with these AGC settings. Although the range data is patchy due to the low texture regions, the traversability map, which is generated over time by merging individual water detection and stereo range images, is dense.

Under RCTA, JPL performed an evaluation of several passive sensors for mud detection, including a Miricle 110KS LWIR camera [13]. Fig. 31 shows a plot of the temperature of dry soil, mud, and air over the entire day, and sample LWIR images from 6:20am, 10:00am, 3:00pm, and 7:00pm. In the temperature plot, the mud was cooler than the dry soil during the entire day, even when there was heavy overcast. The temperature difference between the dry soil and mud was low in the morning, peaked in the mid afternoon, and declined through the evening. The sample imagery confirms that under nominal weather conditions, wet soil is cooler than dry soil throughout the day in TIR imagery. Even at the lowest temperature difference (1.5°C), there was still significant thermal contrast between the dry soil and the mud in the LWIR imagery. Not all the dark regions in TIR imagery are candidate mud, however. Shadows, snow, ice, vegetation, and water can all cause dark regions in TIR imagery. When other sensors are available to disambiguate non-mud dark regions from mud, TIR imagery can provide a strong cue for mud during nominal weather conditions.

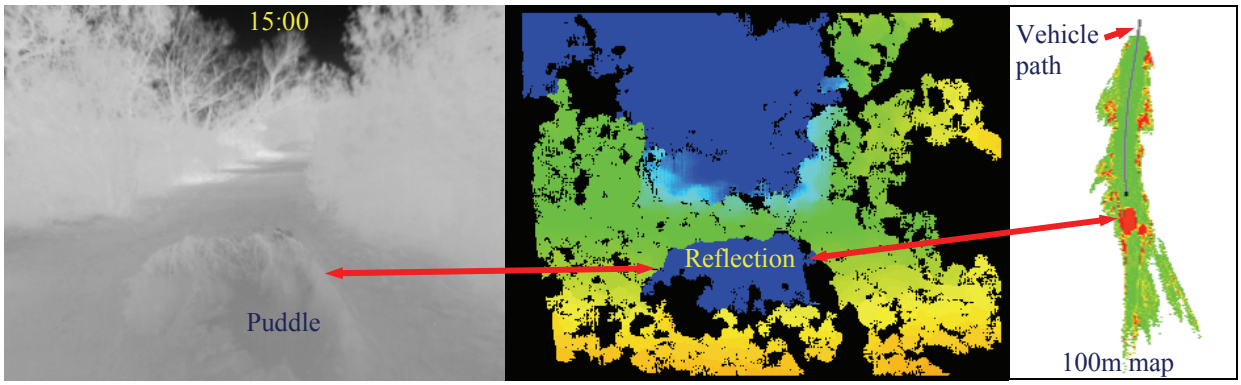


Fig. 30. Still water can be indirectly detected during the day and night by detecting reflections of objects in TIR stereo range data. The range to object reflections is roughly the same as the range to the objects being reflected (middle image). The right image shows a traversability map during a 50 meter southward approach down a dirt trail to the water body. The trees lining the sides of the road and the water body are detected as non-traversable terrain (red). The path of the vehicle down the center of the trail is shown in gray. This sequence was collected at approximately with a stereo pair of Miricle 110KS LWIR cameras.

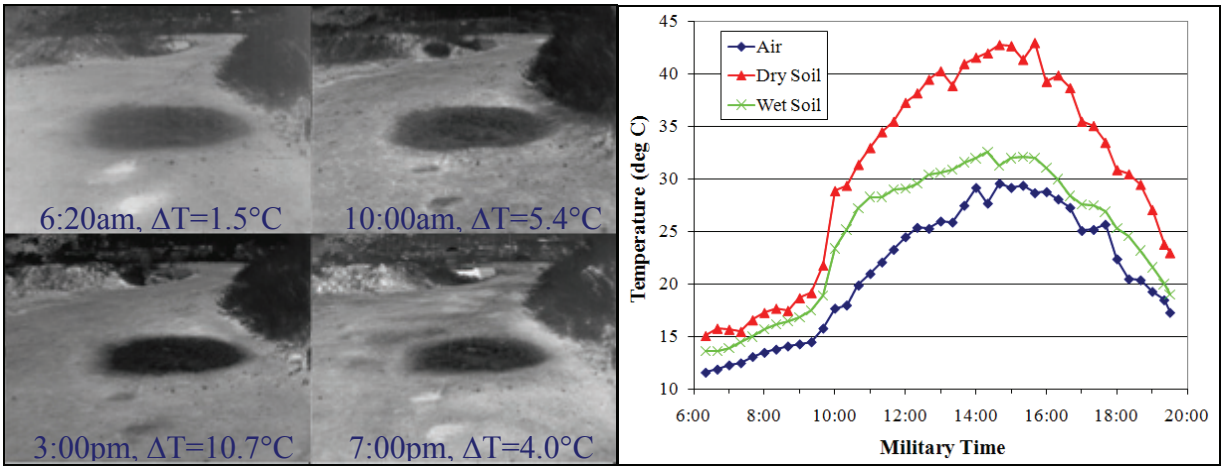


Fig. 31. Under nominal weather conditions, mud has a strong signature in TIR imagery throughout the entire day. ΔT is the difference in temperature between the mud and the surrounding dry soil. This 12 hour data collection was performed with a Miricle 110KS LWIR camera.

7. PEDESTRIAN DETECTION AND CLASSIFICATION

Detecting pedestrians in the vicinity of a UGV is motivated by the requirement to operate safely while navigating urban environments or on primary or secondary roads. In [14], we described our approach to detecting pedestrians using a stereo pair of color cameras. Our approach using a stereo pair of TIR cameras is the same. We generate dense stereo range data, perform visual odometry to compute the change in pose of the UGV, detect and segment regions of interest in the 3D stereo point cloud which have a width and height consistent with an upright human, classify each region of interest with a probability of being human, and then track the regions of interest over multiple frames. Tracking is used to estimate the velocity of objects and reduce incorrect detections by filtering classification probabilities.

Geometric and visual features of each segmented 3D point cloud are used to classify them as human or not human based on shape. The geometric features include the fixed-frame shape moments (variances of point clouds in a fixed frame), rotationally invariant shape moments (the eigenvalues of the point cloud's scatter matrix), and ``soft-counts'' of various width, height, depth, and volume constraints. To compute the visual features, the segmented region of interest is scaled to a canonical size image mask. Visual features are then computed as the moments of a binary, edge, and color foreground mask. Logarithmic and empirical logit transforms of the moments and counts are used to improve the normality of the feature distribution. We then expand the features set to use higher-order terms in a linear classifier to maintain computational efficiency and robustness. Using this feature vector, a Bayesian generalized linear model (GLM) classifier (for logistic regression) is then trained using standard iteratively reweighted least squares (IRLS) to obtain a Gaussian approximation to the posterior mode.

This work is ongoing under the GDRS led Autonomous Navigation System (ANS) program, a component of the U.S. Army's Brigade Combat Team (BCT) Modernization program. Thus far, we have only trained the pedestrian classifier using ground truthed stereo data obtained from color cameras. A preliminary TIR result (using color stereo classification parameters) is shown in Fig. 32 for two frames from the same sequence. This sequence is particularly challenging for stereo ranging because there is low contrast in most of the image. But there is enough contrast on the boundaries of pedestrians to detect and correctly classify them. In the left image, there was no stereo data on the leftmost pedestrian because they were out of the field of view of the right camera. The detected pedestrian on the right side of the images has been tracked over four frames with the vehicle traveling at approximately 15 kph. In the left image, there is stereo range data on the rightmost pedestrian but it is beyond the maximum detection range of our classifier. We are currently in the process of ground truthing TIR stereo imagery and retraining the classifier to obtain classification parameters specific to the TIR sensors. We plan to include visual features when we retrained the classifier using TIR data.

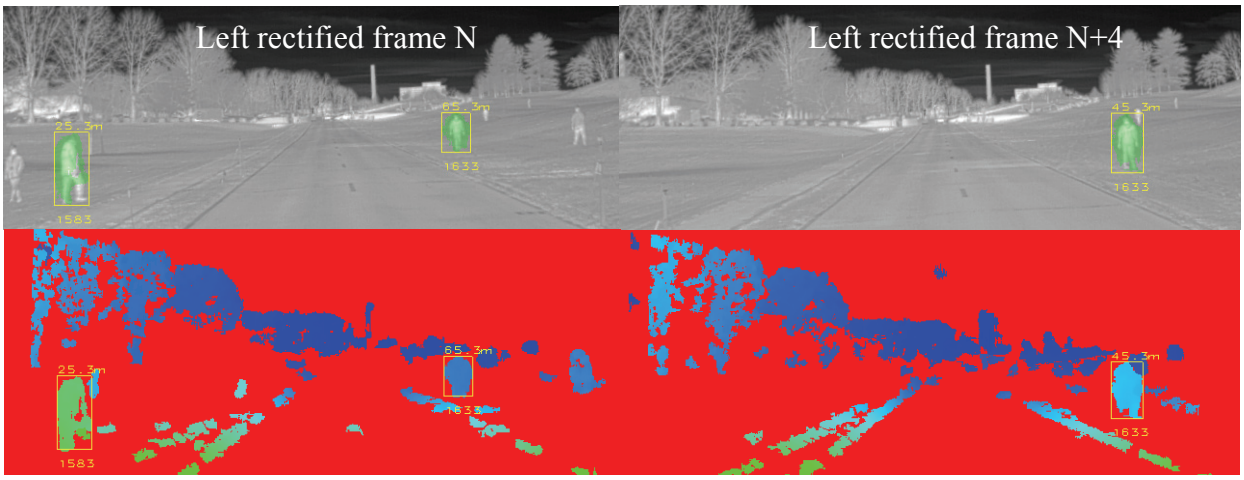


Fig. 32. Two cropped frames of preliminary pedestrian classification results using a stereo pair of Photon LWIR cameras mounted with a 1 meter baseline. The pedestrians within the yellow boxes were detected. The green overlay on the left rectified images (top row) indicates where there was stereo data in the regions of interest detected as pedestrians. The stereo disparity data (bottom row) was generated by processing the images at full resolution with a correlation window of 5x9 pixels. The red pixels indicate where the stereo correlator failed to find a match. The detection range is above each box and a unique tracking identifier is below each box. Visual features were not used for the sequence because the classifier was trained solely on stereo data from color cameras.

8. VEHICLE DETECTION AND CLASSIFICATION

JPL has previously developed a two stage algorithm for finding vehicles in high resolution color stereo images [9]. In the first stage, large 3D blobs in the stereo point cloud are found by projecting the data into a local ground-plane map and looking for peaks in the map. Each detected blob is assigned a feature vector which describes its shape. The blob feature vector is then passed to a linear classifier which determines whether or not the blob is a vehicle. We are currently adapting the algorithm in [9] for use on TIR stereo image pairs during low visibility conditions (night, smoke, fog, etc). A preliminary sample result is shown in Fig. 33 from a dataset of LWIR stereo imagery that was collected at GDRS, Westminster, MD. Because imagery from LWIR sensors typically has lower resolution and contrast than imagery from color sensors, several algorithm parameters require adjustment for optimal use with TIR. We are currently working on improving stereo coverage, blob feature extraction, and classifier methods, and to quantify the results.

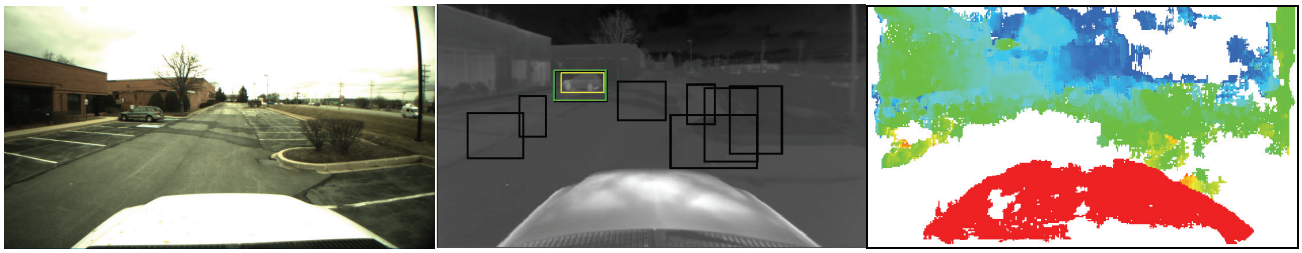


Fig. 33. Preliminary sample result illustrating vehicle detection from a moving vehicle using a stereo pair of LWIR cameras. There are a lot of low texture regions in this scene. To improve the stereo data density, stereo correlation was performed at full resolution (640x480) and half resolution using a 9x15 correlation window and the results were combined. In the middle image, classification results are superimposed on the rectified image from the left camera. The black boxes in the scene identify candidate vehicle blobs that were extracted from the stereo range data. The yellow box indicates manual ground truth annotation, the green box contains a blob correctly identified as a vehicle, and the other black boxes contain blobs correctly classified as not a vehicle.

9. PERCEPTION THRU OBSCURANTS

One benefit of using TIR is its ability to detect emitted energy through fog, dust, rain, snow, and some types of smoke. Over the last decade, JPL and/or GDRS have participated in collecting stereo TIR imagery of scenes containing several types of smoke. JPL and GDRS collected Merlin MWIR stereo imagery through fog-oil smoke during Demo III, JPL collected NC256 MWIR stereo imagery through hexachloroethane (HC) smoke during PerceptOR, and GDRS collected Photon LWIR stereo imagery through smoke generated from a controlled burn of hay bales during RCTA.

Fig. 34 shows an example from the fog-oil sequence where smoke obscures the upper half of the scene containing a vehicle in the monochrome CCD image, but the smoke is only evident in the sky region of the MWIR image. Stereo range data on the ground extends out 15 meters to the visible edge of the ground. The vehicle can be clearly seen in the MWIR stereo range image at a range of 40 meters, but we also get some stereo matching on the smoke itself, hence a floating blob of range data. Fig. 35 shows an example from the HC smoke sequence where smoke heavily obscures the monochrome CCD image, but is not apparent in the MWIR imagery. Stereo range data from the monochrome CCD image is predominantly matching on the smoke, except for the visible road at the bottom of the image. But the stereo range data for the MWIR image does not show effects from the smoke.

Fig. 36 shows an example color and Photon LWIR image from the sequence where smoke was generated from a controlled burn of hay bales. The smoke was generated by breaking rectangular bales of hay into about 0.3 meter segments and arranging these small stacks in rows along a gravel road. A small amount of kerosene was applied to the top of each of the stacks, and then lit, like a giant cigar, but only in the top center of each stack. Once the stacks were burning, each one was folded onto itself so that the burning portion was in the middle of the stack, causing them to burn with little or no flame and lots of smoke. A pedestrian is visible through the smoke in the LWIR image, but not in the color image. Work is ongoing to generate stereo range data for this sequence. But we expect there will dense range data on most of the scene, except where there is image saturation on the hot hay bales and the hottest smoke.

In all three cases, the smoke was considerably more transparent in the TIR imagery, however, wisps of smoke was still visible intermittently in the TIR imagery. It is plausible that a stereo pair of TIR cameras could enable autonomous UGV driving through these types of smoke. However, new algorithm development would be necessary to reason appropriately about stereo matching on the smoke itself.

GDRS has collected Photon LWIR stereo imagery through fog during RCTA. Fig. 37 shows an example color and Photon LWIR image from a sequence driving down a foggy road. A pedestrian is visible through the fog in the LWIR image, but not in the color image. Work is ongoing to generate stereo range data for this sequence. But we expect there will dense range data on most of the scene. A stereo pair of TIR cameras should enable a UGV to drive autonomously through fog.

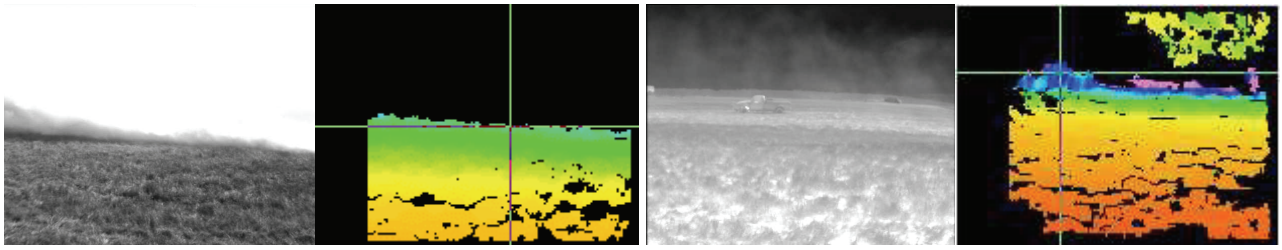


Fig. 34. Monochrome CCD (left) and Merlin MWIR (right) stereo ranging of a scene containing fog-oil smoke. Smoke obscures the upper half of the scene in the monochrome image, but is only evident in MWIR image in the sky region. The stereo range to the cross hairs on the edge of the ground in the monochrome image is 15.04 meters. The stereo range to the cross hairs on the vehicle in the MWIR image is 39.91 meters.

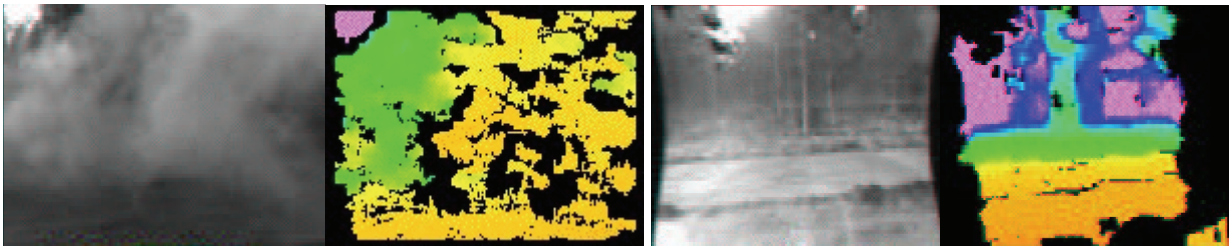


Fig. 35. Monochrome CCD (left) and NC256 MWIR (right) stereo ranging of a scene containing HC smoke. Smoke obscures the most of the scene in the monochrome image, but is not evident in MWIR image. There is stereo matching on the smoke

Fig. 36. Color and Photon LWIR image of a scene containing smoke generated during a controlled burn of hay bales. A pedestrian is obscured by the smoke in the color imagery but is evident in the LWIR image.

Fig. 37. Color and Photon LWIR image of a scene containing fog. A pedestrian is obscured by the fog in the color imagery but is evident in the LWIR image.

10. CONCLUSIONS

For UGV programs that require an ability to perform autonomous navigation during the day and night, and an ability to do it in a stealth mode, TIR cameras can be used to accomplish both requirements. Over the last decade, JPL has developed a variety of perception algorithms under several UGV programs applicable to TIR cameras. These include stereo ranging (in and out of the presence of obscurants), single and dual TIR band terrain classification, negative obstacle detection, tree-trunk detection, pedestrian detection, vehicle detection, and water detection. The quality and resolution of uncooled LWIR cameras have significantly improved over the last decade, making them a viable option for some UGVs with modest speed requirements. The factors that can cause loss of stereo range data with uncooled LWIR cameras include poor camera calibration, image blur from camera horizontal, vertical, and rotational motion, and the presence of low texture regions in a scene. GDRS has designed and fabricated a target for calibrating multiple types of sensors at the same time (including TIR cameras) that overcomes the limitations of previous designs, such as non-uniform heating and limited usability. In UGV experiments with a stereo pair of uncooled LWIR cameras (having a time constant of 7ms), scenes with low-texture regions was more of a stereo ranging limiting factor than image blur when driving on dirt trails at speeds up to 12m/s. One way to reduce the loss of stereo range data in low-texture regions is to perform AGC on a foreground region of interest that excludes the sky. A source of error in the dual TIR band classifier was the approximated registration of images from the two TIR cameras. However, cameras are now available with MWIR and LWIR detectors at each pixel. Use of one of these models is the recommended follow on approach for dual TIR band terrain classification.

ACKNOWLEDGEMENTS

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

REFERENCES

[1] K. Owens and L. Matthies, “Passive Night Vision Sensor Comparison for Unmanned Ground Vehicle Stereo Vision Navigation”, *IEEE Workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications*, Ft. Collins, CO, 59-68, (1999).

[2] K. Hajebi and J. Zelek, “Structure from Infrared Stereo Images”, *Proceedings of the Canadian Conference on Computer and Robot Vision*, 105-112, (2008).

[3] L. Matthies, T. Litwin, K. Owens, A. Rankin, K. Murphy, D. Coombs, J. Gilsinn, T. Hong, S. Legowik, M. Nashman, and B. Yoshimi, "Performance Evaluation of UGV Obstacle Detection with CCD/FLIR Stereo Vision and LADAR", *Proceedings of the IEEE Workshop on Perception for Mobile Agents*, Santa Barbara, CA, June (1998).

[4] P. Bellutta, R. Manduchi, L. Matthies, K. Owens, and A. Rankin, "Terrain Perception for Demo III", *Proceedings of the 2000 Intelligent Vehicles Conference*, Dearborn, MI, 326-331, (2000).

[5] A. Rankin, C. Bergh, S. Goldberg, P. Bellutta, A. Huertas, and L. Matthies, "Passive perception system for day/night autonomous off-road navigation," *Proceedings of SPIE*, Vol. 5804, Orlando, 343-358, (2005).

[6] A. Huertas, L. Matthies, and A. Rankin, "Stereo-Vision Based Tree Traversability Analysis for Autonomous Off-Road Navigation", *Proceedings of the IEEE Work-shop on Applications of Computer Vision*, Breckenridge, CO, (2005).

[7] L. Matthies and A. Rankin, "Negative obstacle detection by thermal signature", *Proceedings of the IEEE Conference on Intelligent Robots and Systems (IROS)*, Las Vegas, (2003).

[8] A. Rankin, A. Huertas, and L. Matthies, "Nighttime negative obstacle detection for off-road autonomous navigation", *Proceedings of the SPIE*, Vol. 6561, Orlando, (2007).

[9] A. Rankin, M. Bajracharya, A. Huertas, A. Howard, B. Moghaddam, S. Brennan, A. Ansar, B. Tang, M. Turmon, L. and Matthies, "Stereo-vision based perception capabilities developed during the Robotics Collaborative Technology Alliances program," Proceedings of SPIE, Vol. 7692, Orlando, April (2010).

[10] A. Rankin, L. Matthies, and P. Bellutta, "Daytime water detection based on sky reflections", *IEEE International Conference on Robotics and Automation (ICRA)*, Shanghai, China, (2011).

[11] A. Rankin and L. Matthies, "Daytime water detection based on color variation," *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Taipei, Taiwan, (2010).

[12] A. Rankin and L. Matthies, "Water detection based on object reflections", submitted to *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, San Francisco, (2011).

[13] A. Rankin and L. Matthies, "Passive sensor evaluation for unmanned ground vehicle mud detection," *Journal of Field Robotics*, 27(4), July/August (2010).

[14] M. Bajracharya, B. Moghaddam, A. Howard, S. Brennan, and L. Matthies, "A fast stereo-based system for detecting and tracking pedestrians from a moving vehicle," *International Journal of Robotics Research*, 28(11-12), (2009).