LWIR passive perception system for stealthy unmanned ground vehicle night operations

Daren Lee*[a], Arturo Rankin[a], Andres Huertas[a], Jeremy Nash[a], Gaurav Ahuja[b], and Larry Matthies[a]
[a]Jet Propulsion Laboratory, California Institute of Technology,
4800 Oak Grove Drive, Pasadena, CA 91109
[b]Space and Naval Warfare Systems Center Pacific (SSC Pacific),
53560 Hull Street, San Diego, CA 92152

ABSTRACT

Resupplying forward-deployed units in rugged terrain in the presence of hostile forces creates a high threat to manned air and ground vehicles. An autonomous unmanned ground vehicle (UGV) capable of navigating stealthily at night in off-road and on-road terrain could significantly increase the safety and success rate of such resupply missions for warfighters. Passive night-time perception of terrain and obstacle features is a vital requirement for such missions. As part of the ONR 30 Autonomy Team, the Jet Propulsion Laboratory developed a passive, low-cost night-time perception system under the ONR Expeditionary Maneuver Warfare and Combating Terrorism Applied Research program. Using a stereo pair of forward looking LWIR uncooled microbolometer cameras, the perception system generates disparity maps using a local window-based stereo correlator to achieve real-time performance while maintaining low power consumption. To overcome the lower signal-to-noise ratio and spatial resolution of LWIR thermal imaging technologies, a series of pre-filters were applied to the input images to increase the image contrast and stereo correlator enhancements were applied to increase the disparity density. To overcome false positives generated by mixed pixels, noisy disparities from repeated textures, and uncertainty in far range measurements, a series of consistency, multi-resolution, and temporal based post-filters were employed to improve the fidelity of the output range measurements. The stereo processing leverages multi-core processors and runs under the Robot Operating System (ROS). The night-time passive perception system was tested and evaluated on fully autonomous testbed ground vehicles at SPAWAR Systems Center Pacific (SSC Pacific) and Marine Corps Base Camp Pendleton, California. This paper describes the challenges, techniques, and experimental results of developing a passive, low-cost perception system for night-time autonomous navigation.

Keywords: Passive perception, long wavelength infrared, stereo vision, autonomous navigation

1. INTRODUCTION

Resupplying forward-deployed units in rugged terrain in the presence of hostile forces creates a high threat to manned air and ground vehicles. An unmanned ground vehicle (UGV) capable of navigating stealthily at night in off-road and on-road terrain for many kilometers from a safe landing zone could significantly increase the safety and success rate of such resupply missions for warfighters. Autonomous UGVs can also increase the mobility and survivability of ground assets during dismounted operations by lightening the load. Passive night-time perception of terrain and obstacle features is a vital requirement for such missions. Under the Office of Naval Research (ONR) Expeditionary Maneuver Warfare and Combating Terrorism Applied Research program, a low-cost perception system with multi-sensor data fusion and passive night-time capabilities was desired to increase the effectiveness of the Marine Corps Air Ground Task Force (MAGTF). As part of the ONR 30 Autonomy Team, the Jet Propulsion Laboratory (JPL) developed a low-cost perception architecture that fuses together data from electro-optical (EO), long wavelength infrared (LWIR), and laser (LIDAR) sensors. This paper focuses on the challenges, techniques, and experimental results of developing a low-cost passive night-time perception capability using stereo vision algorithms with LWIR cameras for autonomous UGV navigation.

2. BACKGROUND

Passive stereo vision algorithms find correspondences in stereo pairs of images to infer depth maps. These depth maps can then be used to create a representation of the surroundings for use in path planning and navigation. LWIR uncooled VOX microbolometer cameras offer low-cost and passive thermal imaging but produce lower image resolutions, higher noise, and longer integration periods than more expensive cooled semiconductor based approaches. The lower image resolutions

*Daren.A.Lee@jpl.nasa.gov

decrease the range accuracy of the stereo vision algorithms and the higher noise decreases the ability of the stereo vision algorithm to find correspondences. Similarly, the longer integration periods of the microbolometer cameras increase artifacts from motion blur, potentially reducing stereo correspondence density. Another challenge in generating accurate range maps using passive stereo techniques is the stereo vision algorithms themselves generate spurious range data from mixed pixels, repeated textures, and low-texture regions. The system requirement challenges include a low-cost requirement that constrains the computing hardware to multi-core CPU-based processing and requires stereo processing at a minimum of 10Hz. The expected operating conditions for the UGV consist mainly of unimproved roads with both man-made and natural structures, with autonomous navigation targeting modest vehicle speeds of a maximum of 5 m/s.

The requirements for the perception system are driven by safe stopping distance and obstacle detection range. We model the stopping distance D on an incline as the summation of the distance required by the frictional forces to decelerate, distance travelled during the time the autonomous system identifies and reacts to a stopping event, and a safe buffer distance:

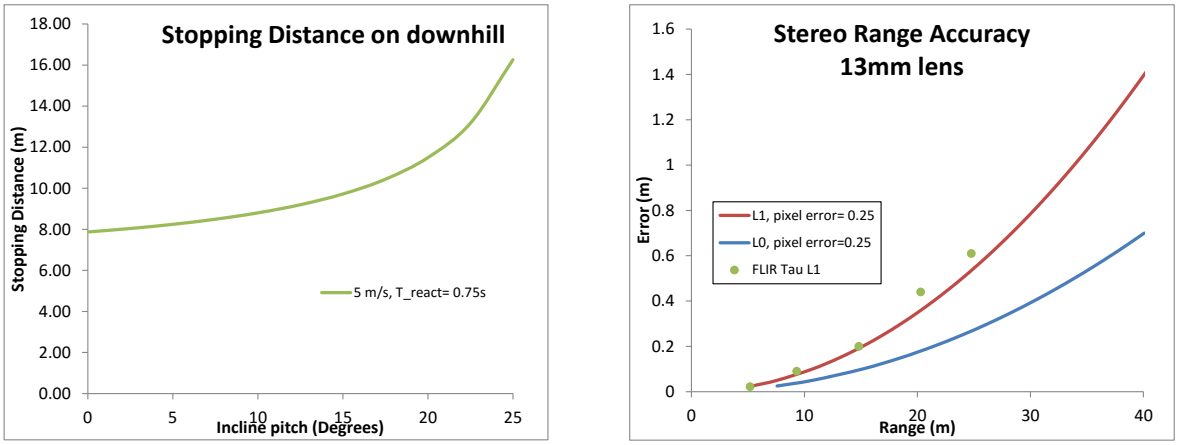
$$D = \frac{v^2}{2g(\mu cos\theta + sin\theta)} + T_{react} * v + D_{buffer} \quad (1)$$

where v is velocity, θ is the incline angle, g is the gravitational constant, μ is the coefficient of friction of the surface, T_{react} is the autonomous system reaction time, and D_{buffer} is a buffer distance. For our expected operating conditions, we use $\mu = 0.6$ for gravel, $T_{react} = 0.5\ seconds$, and $D_{buffer} = 2\ meters$. Figure 1a shows the stopping distances as a function of downhill slope angles at our maximum speed of 5 m/s.

The expected stereo range accuracy is modelled as:

$$\Delta z = \frac{z^2}{fB}\Delta p \quad (2)$$

where Δp is the pixel correlation uncertainty, f is the focal length in pixels and B is the separation baseline of the stereo cameras. Figure 1b shows the theoretical stereo range accuracy for our 13mm focal length thermal cameras with a Δp of a quarter of a pixel and a baseline of 0.75 meters.



(a) (b)

Figure 1. Theoretical models of (a) stopping distance on a sloped downhill at 5 m/s and (b) range accuracy from stereo matching algorithms for 13mm focal length, pixel error of a quarter of a pixel, and a baseline of 0.75 meters at Level 0 (640x512) and Level 1 (320x256). Empirical accuracy results using the FLIR Tau A65 thermal cameras are shown in green.

3. METHODS

The main challenges we faced in the development of our passive LWIR-based night-time perception system were overcoming the low signal-to-noise ratio of the input imagery, increasing the disparity density of the stereo correlator, and removing spurious false positive range data from the depth maps and world model representation. As shown in Figure 2, we apply a series of pre- and post-filters to address these issues in our passive night-time stereo pipeline. These techniques are described below. For the thermal imaging, we use a stereo pair of FLIR Tau A65 thermal cameras with an image

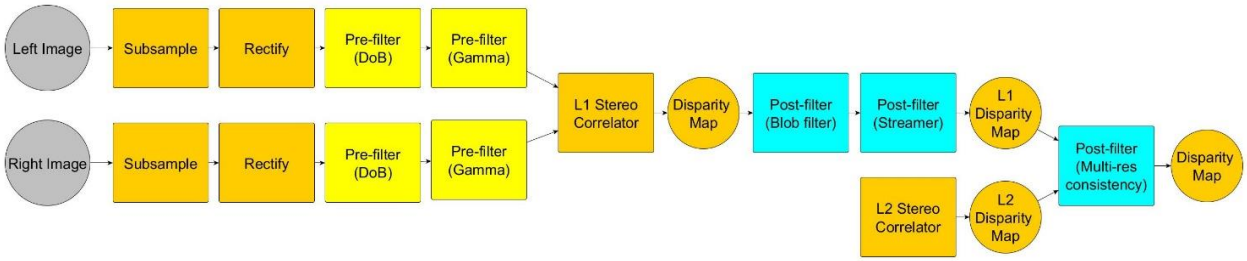


Figure 2. Passive night-time stereo pipeline for LWIR imagery. Applying subsampling to Level 1 (L1) and the pre-filters condition the low SNR imagery for improved disparity density. The post-filters remove spurious disparities to decrease the number of false positive range data. The multi-resolution consistency check compares the disparities at Level 2 (L2) and L1 to remove noisy range data.

resolution of 640x512, an integration time of 10-12 milliseconds, and a baseline of 0.75m. With a 13mm lens, the field of view is 45° horizontal and 37° vertical. Based on empirical data, the effective range data field of view is 33° horizontal.

3.1 Pre-filters

Local, window-based stereo correlators yield improved results with strong texture or features. To condition the imagery without loss of vehicle capabilities, the 640x512 (Level 0) input image are first down sampled by a factor of 2 to 320x256 images (Level 1). While down sampling reduces the inherent noise from the sensor through 2x2 block averaging, it also reduces the range accuracy proportional to the change in image resolution. From our modelling of the stopping distance and expected range accuracy as shown in Figure 1, down sampling to L1 provides detection of obstacles to 20 meters with an expected range error of 0.35 meters, which falls within the stopping distance for various terrain slopes at 5 m/s. Using the lower image resolution for stereo correlation also provides for more computation time for more scene analysis, such as our multi-resolution consistency check we employ to reduce the spurious range data in the disparity maps.

Based on our previous work[1,2], we apply band-pass pre-filtering to reduce the image pair intensity differences and a Laplacian of Gaussian convolution to enhance the underlying image texture through a difference of boxes (DoB) approximation. The normalization of the image intensities is required since approaches that compare local matches based on intensity differences, such as the sum of absolute differences (SAD), are sensitive to contrast changes. Finally, a gamma correction filter is applied to enhance image contrast before matching.

3.2 Stereo Correlation

As in our previous work[2], we leverage the computational efficiency of the SAD similarity metric with SAD5 overlapped correlation windows. The constraints of using only CPU-based processors for stereo processing at 10Hz prohibited the use of algorithms with greater computational complexity, such as semi-global matching. In SAD5, the similarity score of a pixel is given by the SAD score of the center window plus the two best scores of the neighboring four overlapped windows[3]. The neighboring windows improve the correlator along disparity discontinuities by reducing the ambiguity between foreground and background objects. The overlapped windows also create a larger correlation area that helps increase correspondence. To further increase the disparity density, we use a rectangular 11x7 window that is more suitable for outdoor scenes with a receding ground plane. As shown by Kelly and Stentz[4], the optimum aspect ratio of the correlation window is the inverse of the ratio of the corresponding components of the disparity gradient. For outdoor scenes, this translates to a correlation window that is wider than it is high.

3.3 Post-filters

To remove the spurious false positive range data generated by the stereo correlator, we use a blob filter, streamer filter and a multi-resolution consistency check. As in our previous work[2], the blob filter examines the gradient of disparity data to find regions of similar disparity. Regions smaller than a given threshold are removed. The blob filter helps remove speckles of spurious range data. The streamer filter removes artifacts from mixed pixels which overlap both foreground and background depth values. This ambiguity results in a stream of false positive depth values emanating from the foreground object to the background. The streamer filter removes these disparities by examining a trio of range pixels and testing for receding depths above a given threshold.

The multi-resolution consistency check compares the range values at different image pyramid levels to ensure they agree within some threshold. We leverage the coarser and typically denser range values of the lower image resolution to identify false positives in the higher resolution. Our algorithm uses Level 2 (L2) to check the input Level 1 (L1) range data. It first uses the lower resolution as a mask: only the L1 pixels that have valid disparity values in corresponding L2 pixels are checked for consistency; all other L1 depth pixels are invalidated. The multi-resolution consistency check reduces the false positives from repeated textures that often plague local window stereo correlators.

3.4 Probabilistic Sensor Modelling

The uncertainty of range measurements derived from stereo-vision algorithms are known to increase with range. To increase the fidelity of the 3D representation of the environment, our system leverages probabilistic sensor modelling to account for the uncertainty in the range data when creating occupancy grid maps. Occupancy grids are a common approach to integrate multiple sensors into a single obstacle map. Storing occupancy probabilities at each grid cell allows for representations reflecting the uncertainty of sensor properties such as range accuracy. Similar to other work[5,6,7], we use an inverse sensor model to model the occupancy probability to reduce the propagation of false positives into the occupancy grid. We also use ray casting techniques to dynamically update free and occupied voxels as newer data is received.

3.5 Modular architecture

The Robot Operating System (ROS), a message passing based framework, is used as the software platform for the UGV. ROS nodelets enhance performance by eliminating the copy costs when passing messages intraprocess. We leverage ROS nodelets to achieve zero-copy, modular filters that can be interchanged as shown in Figure 3. Since it is the most computationally expensive component, the SAD5 correlator is multithreaded and uses SSE instructions to leverage the multi-core and vectorization capabilities of the target CPUs.

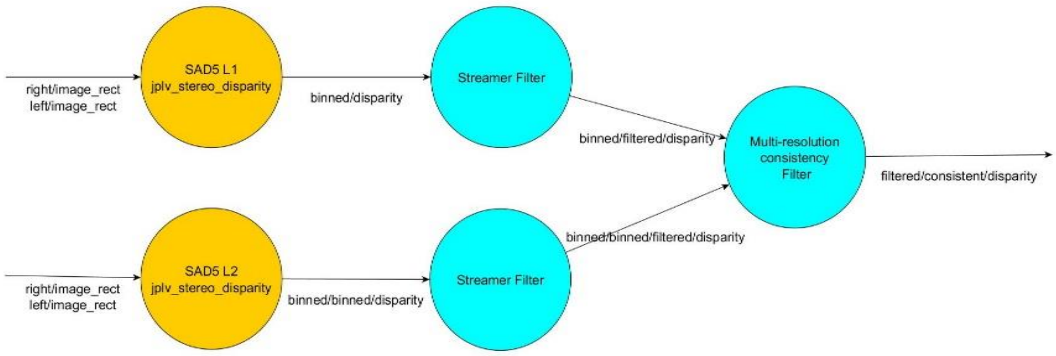


Figure 3. ROS nodelet filters. Each filter is interchangeable and implemented as a ROS nodelet to leverage the zero-copy message passing capability for enhanced performance.

4. RESULTS

The passive night-time perception system was tested at JPL to quantify the stereo matching enhancements, at SPAWAR Systems Center Pacific (SSC Pacific) to quantify the autonomous behavior with our perception system in the loop, and at Marine Corps Base Camp Pendleton to evaluate the UGV in a real world, congested urban environment.

4.1 Stereo Matching Enhancements

Figure 4 shows the impact of our enhancements to the stereo matching results on a scene with unimproved roads using L1 thermal images. The enhancement with the biggest impact on disparity density is applying SAD5 versus SAD1. With the overlapping windows, the near and far fields show an increase in disparity values. Switching to a rectangular 11x7 correspondence window generally increases the smoothness of the disparity values. Increasing the gamma, on the other hand, is more scene dependent as it helps increase the disparity density only in areas of low contrast, such as in the far field in Figure 4c. As shown in Figure 4b and Figure 4c, the disparity density decreases for on-the-move data but with our enhancements, the high priority areas of the unimproved road are clearly identified at our max speed.

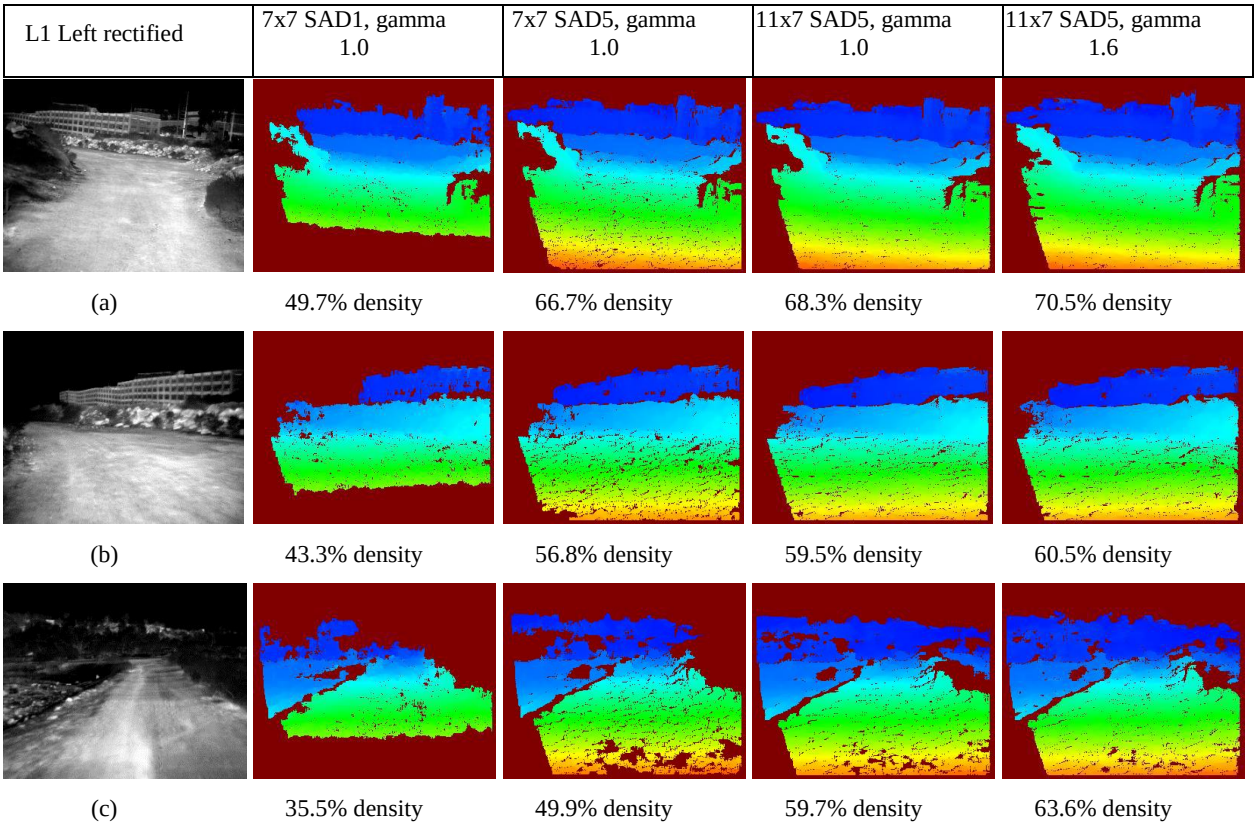


Figure 4. Stereo matching comparison with different configurations on a scene with unimproved roads with our passive LWIR night-time stereo pipeline with (a) static scene, (b) on-the-move near static scene, and (c) at peak speed of 4.87 m/s. The density is the percentage of pixels in the entire image with disparity values. SAD5 yields an increase in density at the near and far fields compared to SAD1. The 11x7 window increases the smoothness and the gamma helps improve the density in low contrast areas.

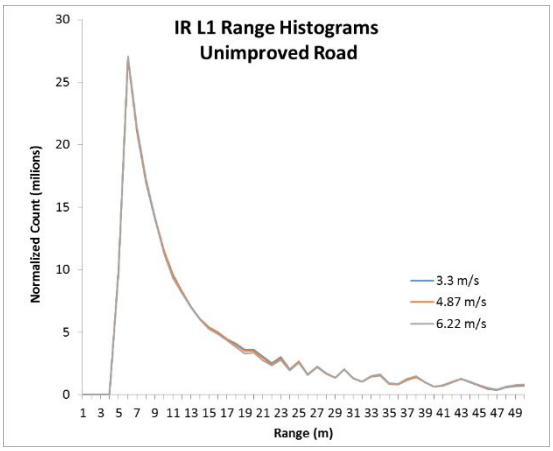
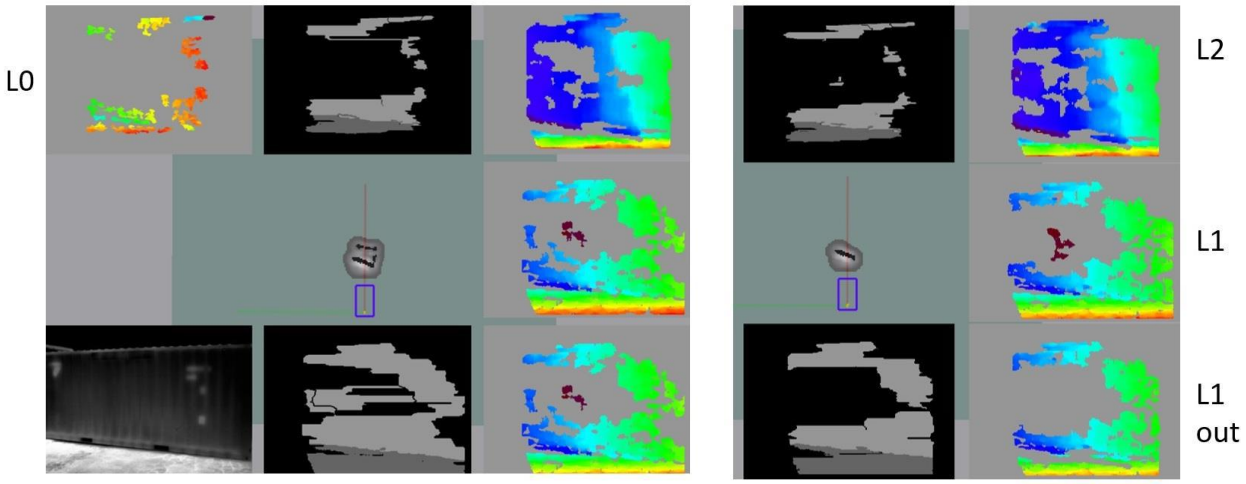


Figure 5. Comparison of range histograms at three different vehicle speeds on the same course to quantify the effects of forward motion blur. At these target speeds, the motion blur does not significantly cause disparity drop outs.



(a) (b)

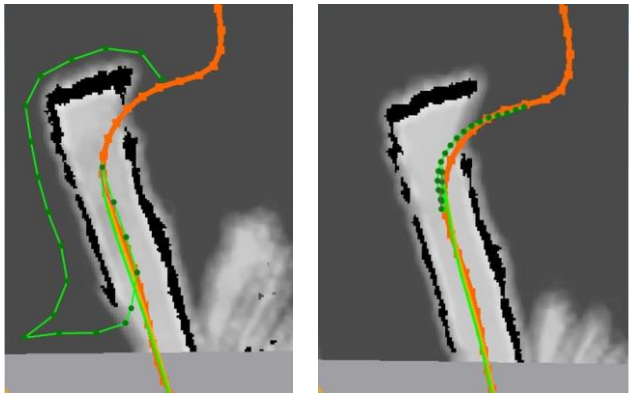
Figure 6. Multi-resolution consistency check uses the coarser but typically denser range data at a lower resolution to remove false positives. The lower left image in (a) shows a container wall with a repeated pattern that causes ambiguity in the correlator and a false wall to appear behind the actual wall. The consistency check in (a) preserves all values in L1 and only performs the consistency check if there exists corresponding range data in L2. This leads to the preservation of the false positive range data. In (b), L2 is used as a mask for the consistency check. Only those values in L1 with corresponding valid depths in L2 have the consistency checked performed. All other pixels in L1 are invalidated, resulting in the removal of the spurious range data behind the container wall.

With the longer integration times of the uncooled microbolometer sensor, the disparity density may be reduced due to the pixel blur from motion. To test the impact of forward motion blur artifacts on our passive night-time stereo pipeline, the vehicle was manually driven at three different speeds on the same course with unimproved roads. To compare the results, a histogram of range values normalized by frame count was computed for each run. With strong motion blur, we would expect disparity drop outs closer to the vehicle since the pixel blur from motion would be greater at these ranges. As shown in Figure 5, for the modest speeds tested, motion blur did not significantly cause any disparity drop outs.

4.2 Post-filters

Figure 6 shows the impact of the multi-resolution consistency check filter. A container with repeated texture causes ambiguity in the stereo correlator and a false wall to appear behind the container wall. In Figure 6a, the multi-resolution consistency check algorithm preserves all L1 data and only applies the consistency check if there exists corresponding data in L2. While this approach yields higher density, it also preserves possible spurious data. In Figure 6b, the multi-resolution consistency check algorithm uses L2 as a mask and the consistency check is only applied to pixels in L1 with corresponding valid depths in L2. In our example, the spurious range data behind the container wall with repeated textures is removed. This approach relies on the coarser L2 disparity having broad density coverage, which is typically the case with the lower resolutions. In the cases where there are gaps in the L2 data, this approach can potentially remove good pixels.

Figure 7 illustrates the impact of the post-filters in reducing the false positives in an L-shaped test apparatus. Without the post-filters, the far wall appears thicker at the exit. This causes the autonomous system to plan a reverse route to circumvent the perceived closed opening. With the post-filters, the false positives are removed, allowing the autonomous system to plan a route as expected.



(a) (b)

Figure 7. Impact of the post-filters to remove false positive range data on world model cost map. The black areas are obstacles, the white areas are clear areas without obstacles, and the dark grey are unknown areas. The orange line represents the target input navigation path and the green line represents the planned path by the autonomous system. Without the post-filters in (a), the autonomous system detects a false obstacle and reroutes to circumvent the blockage. With the post-filters in (b), the spurious false positive are removed and the autonomous system plans as expected.

4.3 Constrained environment performance

To evaluate the night-time autonomous performance in a real world environment, test trials were carried out in a realistic, constrained urban course with building structures, unimproved roads, small and large vegetation, and barrel obstacles. Each segment of the course stressed different components of the autonomous system. Human intervention counts for safety stops and planner timeouts were collected for each trial segment. Safety stops are interventions from the human safety driver to stop the autonomous vehicle due to safety concerns from system performance. Planner timeouts occur when the autonomous vehicle stops moving and is unable to plan a suitable path for a period of 5 seconds. The UGV with our passive night-time perception system was tested an hour after sunset. To mitigate navigation issues with the limited field of view, the planner was configured to be more conservative by increasing the cost for navigating into unknown areas.

For the 17 segments tested, the UGV successfully navigated 8 of the segments without any interventions, 7 segments were halted due to safety stops, and 2 segments were halted due to planner timeouts. The mitigations we employed to reduce the spurious false positives yielded accurate obstacle detections throughout the course. Barrels were detected at 15-18 meters giving the navigation system suitable distance to plan appropriate maneuvers. The majority of building structure footprints were detected without spurious data from repeated textures of the wall studs.

In the segments where the obstacles were clearly in the field of view of the thermal cameras, the UGV performed well. In the segments where a wider field of view of the environment was necessary to navigate successfully, false negatives were the major cause of safety stops. Corridors slightly wider than the narrow field of view and tight corners caused missing barriers to appear in the system, yielding undesirable route planning or latent obstacle detection.

5. CONCLUSION

We have described our passive LWIR-based night-time perception system that overcomes the low signal-to-noise ratio of the input imagery, increases the disparity density of the stereo correlator, and removes spurious false positive range data from the depth maps through a series of pre- and post-filters, stereo correlator enhancements, and probabilistic sensor modelling. Our approach reduces the false positive range data to generate suitable representations of the environment for night-time UGV navigation. Terrain and obstacles are detected at suitable distances for maneuver decisions and the spurious false positives associated with mixed pixels and repeated textures are greatly reduced. With the limited field of view, however, false negatives are more prevalent which inhibits successful autonomy. Improving the field of view is the focus of future work and could be addressed by using a shorter focal length lens, modifying the navigation software to drive for perception before entering unknown areas, or moving toward a pan-tilt camera system.

ACKNOWLEDGEMENTS

The research described in this paper was carried out by the Jet Propulsion Laboratory, California Institute of Technology, and was sponsored by ONR, 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. The authors would like to thank John Andrews, Michael Bruch, Chris Scrapper, Ryan Halterman, Adam Nans, Greg Kogut, Josh Zapf, Thomas Denewiler, Mark Tjersland, Donnie Fellars, Greg Droge, Jacoby Larson, and Saam Ostovari.

REFERENCES

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

[2] Rankin, A., Bajracharya, M., Huertas, A., Howard, A., Moghaddam, B., Brennan, S., & Matthies, L., "Stereo-vision based perception capabilities developed during the Robotics Collaborative Technology Alliances program", *Proceedings of SPIE*, Vol. 7692, pp. 76920C1-76920C15, (2010).

[3] Hirshmueller, H., "Improvements in real-time correlation-based stereo vision," *IEEE Conf. Computer Vision and Pattern Recognition*, pp. 141–148, (2001).

[4] Kelly, A., and Stentz, A., "Stereo vision enhancements for low-cost outdoor autonomous vehicles," *Int'l Conf. on Robotics and Automation, Workshop WS-7, Navigation of Outdoor Autonomous Vehicles*, Vol. 1, (1998).

[5] Matthies, L. and Elfes, A., "Integration of sonar and stereo range data using a grid-based representation," *IEEE International Conference on Robotics and Automation*, pp. 727-733, (1988).

[6] Andert, F., "Drawing stereo disparity images into occupancy grids: measurement model and fast implementation," *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 5191-5197, (2009).

[7] Heng, L., Honegger, D., Lee, G. H., Meier, L., Tanskanen, P., Fraundorfer, F., & Pollefeys, M., "Autonomous visual mapping and exploration with a micro aerial vehicle," Journal of Field Robotics, 31(4), 654-675, (2014).