

ADVANCED PERCEPTION SYSTEM FOR VEHICLE AUTONOMY

William C. Flannigan
Michael P. Rigney, Ph.D.
Kevin J. Alley
Southwest Research Institute
San Antonio, TX

ABSTRACT

An approach for a perception system for autonomous vehicle navigation is presented. The approach relies on low-cost electro-optical (EO) sensors for terrain classification, 3D environment modeling, and object/obstacle recognition. Stereo vision is used to generate real-time range maps which are populated into a hybrid probabilistic environment model. Textural and spectral cues are utilized for terrain classification and spatial contextual knowledge is proposed to augment object recognition performance.

INTRODUCTION

Current sensor suites demonstrated on autonomous vehicles are prohibitively expensive for mass deployment. In addition, current sensors (specifically LIDAR) provide obstacle detection, but lack robust object classification or perception capabilities sufficient for advanced decision making. An autonomous vehicle needs the ability to determine, for example, if an obstacle can be driven over/through, as in the case of a tall grass, brush, or a wooden barricade, or if it needs to circumnavigate the obstacle, as in the case of a concrete or rock barrier.

Autonomous vehicles require enhanced sensing capability to fulfill current and future mission scenarios. Dismounted troops have virtually unlimited mobility. An autonomous support vehicle will therefore need to traverse any terrain within the operational envelope of the vehicle. This could include off-road terrain, grades, rough surfaces (rocky), vegetated (overgrown) paths, dirt roads, narrow paths, and stream/drainage (wet/dry) crossings.

The Advanced Perception System (APS) is being developed to demonstrate enhanced perception capabilities utilizing low-cost sensors (electro-optical cameras). The program seeks to affirm the hypothesis that advanced analysis algorithms can be applied to data from economical sensors to achieve advanced perception capability. The initial program goal is to demonstrate obstacle/object/terrain classification providing data for use by path planning algorithms that is superior to that provided by current LIDAR systems. The sensor and data processing approach will parallel the human ability to achieve a high level of

perception from only visual data. Specifically, the proposed solution will integrate:

- Texture analysis of Electro-optical (EO) imagery for object/region segmentation and classification
- Multispectral image analysis for object/region segmentation and classification
- Range estimation from EO imagery for region segmentation, object size and pose estimation
- Tiered classifiers coupled with context based reasoning for obstacle/object perception

System performance will be evaluated and demonstrated in off-road and low-use dirt road environments. The program is sponsored by the Office of Naval Research, Code 30 in support of Marine Corps Enhanced Company Operations and the Small Unit Mobility Enhancement Technology (SUMET) Program. This paper describes the high-level system architectures and preliminary results.

PROGRAM OBJECTIVES

The APS development is part of a multi-year program to evaluate unmanned system technologies for small and distributed units. The APS is envisioned to provide sensory input for higher level path planning and cognitive reasoning systems. Several system-level objectives have been identified:

- **Terrain Classification** The APS should provide off-road terrain classification sufficient for traversability

analysis and path planning activities for varied mission plans and risk states.

- **Object Classification** The APS will require object classification for objects of interest, such as pedestrians and vehicles, that are important to high level behavior generation.
- **Modularity** The APS is platform agnostic and should provide interfaces that enable easy adaptation to varied hardware systems and missions.
- **Low Cost** A hardware (sensors, computational resources, communications) replication cost target of less than \$30,000 has been identified to enable high volume application of the technology.

SYSTEM ARCHITECTURE

The APS hardware is focused on commercial-off-the-shelf (COTS) EO sensors in the visible and near-infrared spectrum. Future efforts will integrate other vehicle sensors such as GPS/IMU, wheel speed sensors etc. Current development is being conducted on Southwest Research Institute's (SwRI's) Mobile Autonomous Technology Research Initiative (MARTI) platform [1], shown in Figure 1, and will be transitioned to a militarily relevant vehicle for future demonstration milestones. A pair of color, 1.4 Megapixel, Gigabit Ethernet cameras is mounted to the vehicle roofline. These imagers are used for stereo disparity analysis for range calculation and for textural cues. A second multispectral camera is employed for color classification. Keeping with the low-cost COTS goals, the multispectral camera is composed of several (up to six) 0.36 Megapixel monochromatic imagers with individual optical filters, mounted in a compact assembly. Figure 2 shows a high-level conceptual architecture for the APS.



Figure 1: SwRI's MARTI Autonomous Vehicle

The software architecture makes use of the Robot Operating System (ROS) which is actively developed and

maintained by Willow Garage [2]. ROS is an open-source framework for inter-process communication for robotic and sensor systems. ROS provides an architecture that fosters modularity and distributed processing which will be important as the APS becomes more sophisticated and is interfaced with other autonomy and hardware modules.

The APS is executed on general purpose CPUs either resident within the MARTI vehicle or on portable systems. Future efforts will optimize the computational architecture for size, weight, and power consumption. Real-time performance for control of the autonomous platform is not yet required, and computational performance and future requirements are being characterized in the initial development phase.

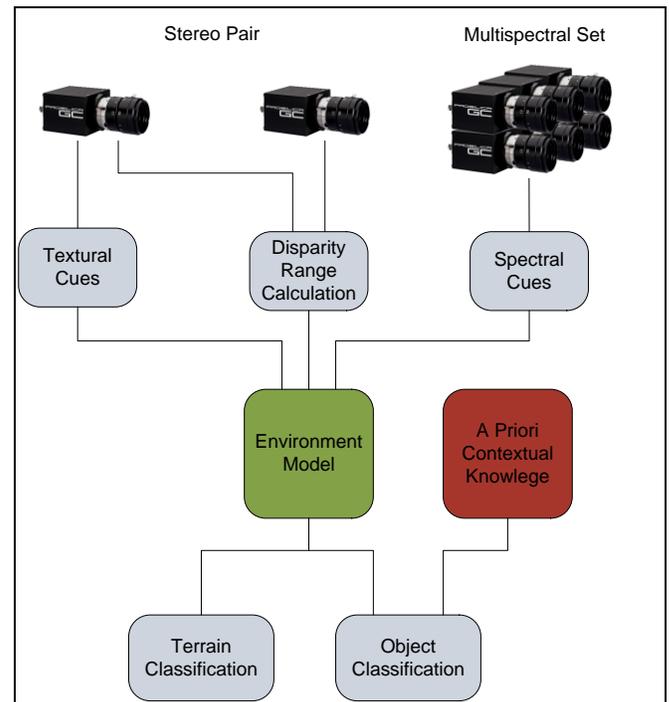


Figure 2: High Level System Architecture

STEREO VISION

Stereo imaging has several potential advantages over other sensing modalities for autonomous navigation. It is passive, requiring only ambient light for sensing. It is relatively low cost with high quality commercial-grade imaging sensors. There are no moving parts, improving reliability. And, stereo vision can provide rich spatial, textural, and color information for sophisticated perception capabilities. Indeed, biological systems perform incredibly complex behaviors using vision sensing alone.

Stereo image processing, however, has drawbacks which have limited its use on many autonomous systems to date. It is computationally expensive, although modern multi core

parallel processing architectures have significantly improved performance. It requires sufficient ambient light (or artificial illumination) for operation. Analysis of the resulting 3D image data is complex compared to object detection algorithms used for analyzing LIDAR data, for example, which typically provides relatively limited scene coverage.

Much of the APS' ability to characterize the environment relies on accurate, high-speed, and robust stereo mapping capability. Several open-source stereo processing pipelines have been evaluated and the Semi-Global Block Matching (SGBM) algorithm described by Hirschmüller [3] has been implemented. Our experience indicates that SGBM provides a good compromise between computational speed and disparity map quality. In addition, it performs well under varied image intensities. Finally, its complexity is linear with respect to pixel count and disparity range, permitting exploration of wide stereo baseline configurations and high resolution imaging.

A pair of 1.4 MP cameras is mounted on a bar at the vehicle roofline, permitting adjustments to the stereo baseline. In general, increasing the baseline provides potential improvement in range accuracy but imposes greater computational demands because of increased disparity window size, and yields increased noise due to poorer image correlation. In addition, image resampling tools have been implemented which permit investigation of varied image resolutions. Increased image resolution can capture finer detail and potentially improve long range performance, again at the cost of computational complexity.

Figure 3 shows the processing rate at varied resolutions and at two baselines on a Core i7 mobile processor. The speed variation at the two baselines is due to an increased disparity window. High update rates (~10 Hz) will be required for real-time near-field obstacle avoidance, while high resolution is desired for mid to far field sensing. Future work may implement separate stereo processing pipelines dedicated to near field object detection and far field path planning. Each pipeline could operate at different resolutions and with potentially different algorithms optimized for each task. Alternately, one could take a pyramidal approach as described by Rankin et al. [4]

Figure 4 shows the range error from an experiment where the MARTI vehicle's LIDAR sensors were used as ground truth range measurements while the vehicle was driven towards a set of barriers that provided a flat vertical target. The LIDAR distance to the barriers was averaged from the multiple returns and the stereo point cloud was fit to a plane. The graph shows the difference between the two measurements. The disparity analysis relies on intrinsic and extrinsic camera calibration parameters which directly affect the range accuracy. For these experiments, the cameras were calibrated with targets between 3 and 10 meters from

the cameras. New calibrations will be investigated to determine if longer range calibration targets can improve the long range accuracy.

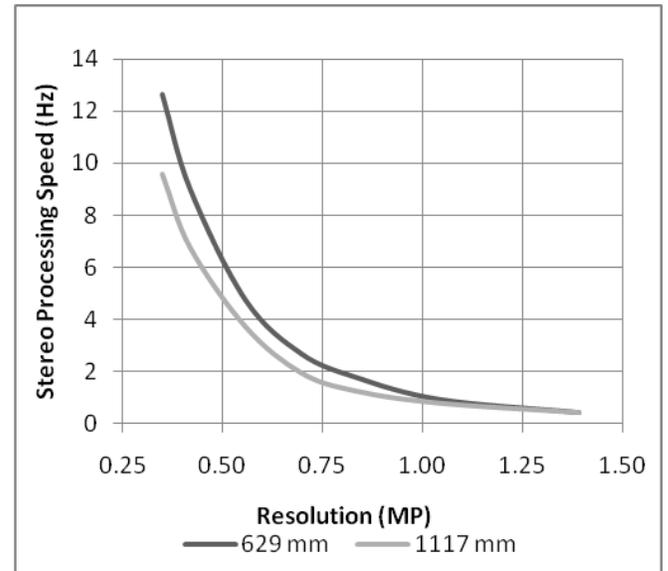


Figure 3: Stereo Processing Rate vs. Resolution at Two Baselines on an Intel Core i7-720QM Processor

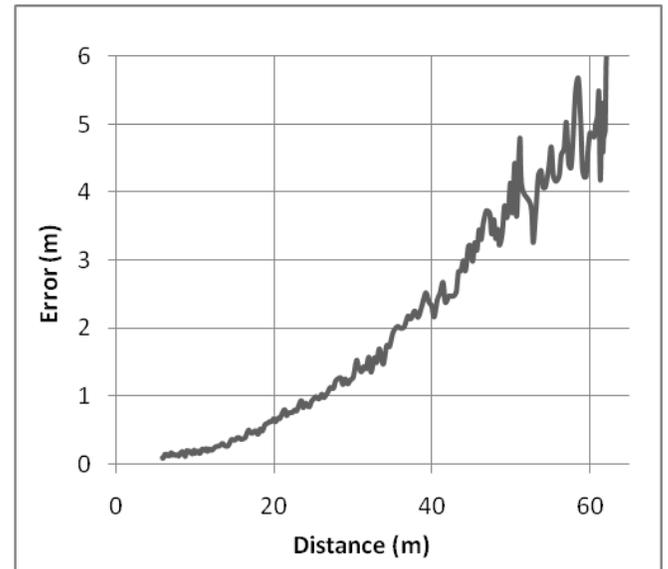


Figure 4: Range Error Using 1117mm Baseline and 680x512 Image Resolution

ENVIRONMENT MODEL

The raw point clouds generated from the stereo vision are not readily useful for path planning and obstacle avoidance; they tend to be noisy and large, requiring excessive memory and storage for anything but small maps. Instead, a compact

environment representation is needed that has the following features:

- **Compact** The model needs to provide a reasonably compact representation for both RAM and long-term storage.
- **Probabilistic** Due to the noisy, high resolution nature of the raw point clouds, the environment model should be probabilistic in nature, providing a statistical representation of the data.
- **Hierarchical/Scalable** Near field navigation may require high resolution representations, but far-field route planning can be much more efficient with low resolution data. The environment model should permit easy access to the data at varied resolutions.
- **Extensible** The model should be dynamically expandable at run-time so that it can incorporate new data as the vehicle progresses.
- **Flexible** Due to the nature of our analysis, the model will be called upon to represent data types beyond simple occupancy. It will need to encapsulate class information, confidence levels, object identifiers, color, and geometric properties.

To accommodate these needs, we have adopted a hybrid representation of the ground surface and the objects that reside on it. Our obstacles model is based on the OctoMap presented by Wurm et al. [5]. We have extended Wurm's approach to include varied data beyond simple occupancy. The model is a voxel-based octree where each leaf node represents a volume of 3D space. Each level in the tree provides a convenient representation of the space at a nominal resolution. Figure 5 shows an example of a tree at three different resolutions.

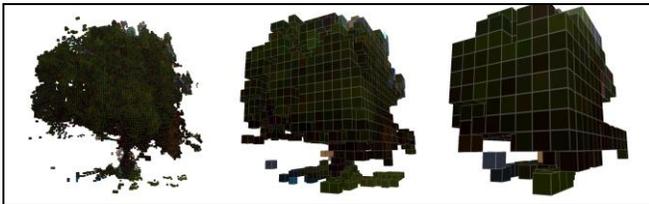


Figure 5: Variable Resolution Representations of a Tree

The ground is extracted from the point cloud via the Random Sample and Consensus (RANSAC) method [6] for plane fitting. The ground model is represented in a quadtree which is analogous to the octree used for the objects, but yields a more compact and precise surface representation. The ground model also enables easy terrain topology calculation, shown in Figure 6. The local slope of the terrain will be an important factor in calculating traversability costs.

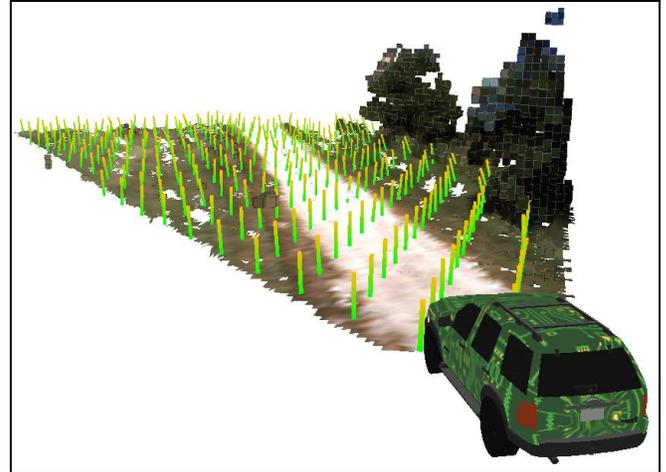


Figure 6: Full Environment Visualization with Lines Representing Ground Surface Normal Vectors

Since the ground in off-road environments does not necessarily fit to a plane well, a multi-segment approach has been implemented to account for variations in terrain topology. The current implementation uses three segments separated by their distance from the vehicle: 0-15 meters, 15-30 meters, and 30 meters and beyond. This approach permits classification of the ground in terrain with moderately changing slope.

The confidence of the data at any location in either the ground or obstacle models is a function of the number of samples within a region and the distance of the region from the vehicle. The current model applies a confidence value that is inversely proportional to the distance from the vehicle to the sixth power. This exponential weighting gives a high value to near-field readings which helps accommodate moving objects and inaccuracies in range estimation at large distances. The confidence value at any region is the weighted average of the existing values and the new reading.

TEXTURE ANALYSIS

Texture analysis is applied to images from one of the high-resolution stereo cameras. Several texture filters are being evaluated for efficacy for region segmentation and classification. The texture filters operate in the 2D image space which is then directly mapped to the 3D environment model. Initial efforts have focused on 2D Gabor filters [7]. The Gabor kernels are dependent on factors of scale, frequency, and orientation providing high dimensionality feature spaces. Principle Component Analysis (PCA) is used to provide the most significant features [8]. The results of the PCA will ultimately be used as part of the feature vector for a terrain classifier. Initially, k-means clustering is used to visualize the performance of the filters, as shown in Figure 7.

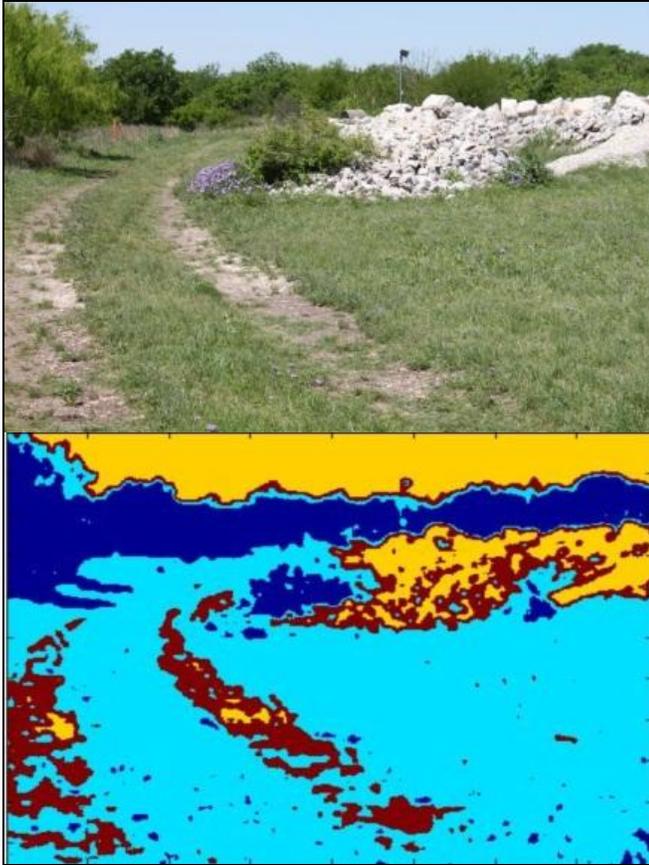


Figure 7: A four class K-means Cluster Based on Gabor Texture Filters

SPECTRAL ANALYSIS

As part of an internally funded research project, SwRI investigated the selection and use of optimal multispectral images for terrain classification. Hyperspectral reflectance data (120 bands) from materials of interest was analyzed to select a small set of commercially available spectral filters yielding maximum material classification performance. Performance obtained from multispectral filter sets was compared to that obtained using the full hyperspectral data set. Results showed that 90% of the classification performance (relative to hyperspectral) could be achieved using only six spectral bands obtainable from COTS optical filters [9]. This approach is being ported to the APS to provide a low cost multispectral imager that has comparable performance to a high-band hyperspectral camera. Figure 8 shows the result of a terrain classification using the multispectral approach.

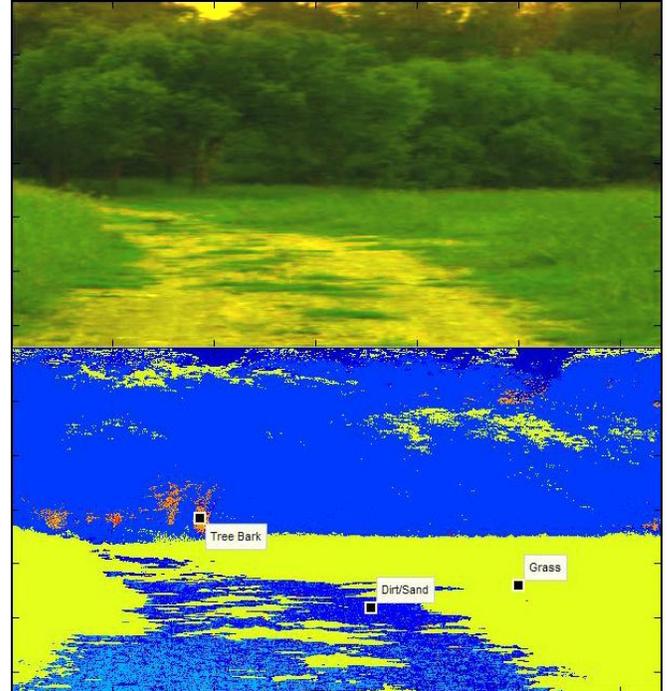


Figure 8: Automated Terrain Classification Based on Multispectral Analysis

CLASSIFICATION

A requirements analysis has led to the development of a number of material classes relevant to autonomous navigation in a primarily off-road environment. The current effort will use supervised training to implement classifiers for identification of the material classes shown in Table 1. Both textural and spectral features will be inputs to the classifier. Material classification output will be accumulated in the environmental model, where probabilistic reasoning for surface/object recognition can be applied. The material classes were selected based on their impact on traversability in off-road environments. For example, grass in certain scenarios and missions may be drivable and in others would be assigned a high traversability cost.

Table 1: Material Classes

Wood	Dirt	Rock/Concrete
Asphalt	Grass	Foliage
Sky	Water	Other Obstacle

FUTURE WORK

The APS is in its first year of development and an initial milestone demonstration will be performed within the calendar year. Prior to this demonstration, the texture and spectral filters will be fully implemented and used as input to the classifier(s). The demonstration will provide the

opportunity to evaluate the classification performance and characterize the computational requirements for the system.

Tasks outlined for next year include performance optimization and hardware development for a real-time system capable of interfacing with the vehicle autonomy system. In addition, the spatial contextual reasoning will be implemented to begin efforts on militarily relevant object detection/recognition.

REFERENCES

- [1] <http://marti.swri.org/>
- [2] M. Quigley, B. Gerkey, K. Conley, J. Faust, T. Foote, J. Leibs, E. Berger, R. Wheeler, and A. Ng, ROS: an Open-source Robot Operating System, ICRA Workshop on Open Source Software in Robotics, 2009.
- [3] H. Hirschmüller, “Stereo Processing by Semi-Global Matching and Mutual Information”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 30(2), February 2008, Pages 328-341, 2008.
- [4] A. Rankin, M. Bajracharya, A. Huertas, A. Howard, B. Moghaddam, S. Brennan, A. Ansar, B. Tang, M. Turmon, and L. Matthies, “Stereo-vision based perception capabilities developed during the Robotics Collaborative Technology Alliances program”, Proceedings of the SPIE, Volume 7692, 2010.
- [5] K. Wurm, A. Hornung, M. Bennewitz C. Stachniss and W. Burgard, “OctoMap: A Probabilistic, Flexible, and Compact 3D Map Representation for Robotic Systems”, Proc. of the ICRA 2010 Workshop on Best Practice in 3D Perception and Modeling for Mobile Manipulation, 2010.
- [6] M. Fischler and R. Bolles “Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography”, Comm. Of the ACM, Volume 24, Pages 381-395, 1981.
- [7] M. Porat and Y. Zeevi, “The Generalized Gabor Scheme of Image Representation in Biological and Machine Vision,” IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 10, Number 4, Pages 452-468, 1988.
- [8] H. Lu, Y. Liu, Z. Sun and Y. Chen, “An Active Contours Method Based on Intensity and Reduced Gabor Features for Texture Segmentation”, IEEE international conference on Image processing, Pages 1361-1364, 2009.
- [9] SwRI IR&D Project 09-R8081 Final Report – Hyperspectral Band Reduction Ground Perspective Applications.