Vegetation Versus Man-Made Object Detection from Imagery for Unmanned Vehicles in Off-Road Environments

Josh Harguess and Jacoby Larson

Space and Naval Warfare Systems Center Pacific, 53560 Hull St., San Diego, CA, USA

ABSTRACT

There have been several major advances in autonomous navigation for unmanned ground vehicles in controlled urban environments in recent years. However, off-road environments still pose several perception and classification challenges. This paper addresses two of decisions must be made about obstacles in the vehicle's path. The most common obstacle is vegetation, but some vegetation may be traversable, depending on the size of the vehicle and the type of vegetation. However, man-made objects should generally be detected and avoided in navigation. We present recent research towards the goal of vegetation and man-made object detection in the visible spectrum. First, we look at a state-of-the-art approaches to image segmentation and image saliency using natural scene statistics. Then we apply recent work in multi-class image labeling to several images taken from a small unmanned ground vehicle (UGV). This work will attempt to highlight the recent advances and challenges that lie ahead in the ultimate goal of vegetation and man-made object detection and classification in the visual spectrum from UGV.

Keywords: vegetation detection, vegetation classification, man-made object detection, unmanned ground vehicle, segmentation, image labeling, visual saliency, natural scene statistics

1. INTRODUCTION

Research into the area of autonomous navigation for unmanned ground vehicles (UGV) has accelerated in recent years. This is partly due to the success of programs such as the DARPA Grand Challenge[1] and the dream of driverless cars,[2] but is also due to the great number of challenges that lie ahead in making this dream a possibility. The Department of Defense (DoD) is interested in researching solutions for the various challenges of UGVs for numerous applications and in many different environments, including rough off-road terrain.

In these off-road environments, there are several perception and classification challenges that remain. Many previous researchers have addressed the segmentation and classification of imagery to identify road and non-road regions for navigation as well as obstacles in the vehicle's path. The most common obstruction in off-road environments is vegetation, which may or may not be an obstacle depending on the vehicle. The desired solution would detect and classify vegetation, then measure its compressibility and porosity to determine traversability of the UGV through the detected region. Also, detecting and classifying man-made objects in the same perceptual system is imperative so that the vehicle may avoid such obstacles if necessary.

In this work we examine the problem of detecting and classifying vegetation and man-made objects from imagery. The strict power, size, and weight payload constraints of a small UGV move this research toward a camera-only solution. Also, by using only a camera, the perception is passive and not active.

We present the results of recent research towards the goal of vegetation and man-made object detection in the visible spectrum. The paper is organized as follows. In Section 2, we review the background literature for obstacle detection and navigation for UGVs as well as related research towards the goal of vegetation and man-made object detection. In Section 3 we introduce our problem of vegetation and man-made object detection from a small UGV in an off-road environment. We apply state-of-the-art methods in image segmentation, saliency and multi-class image labeling to four representative images for qualitative analysis. We conclude the paper in Section 4 with closing thoughts and a look to future work in this area.

Please send correspondence to Josh Harguess:
E-mail: joshua.harguess@navy.mil, Telephone: 1 619 553 0777

2. RELATED WORK

Research in the area of autonomous vehicle navigation is vast, covering sensors of various types, approaches from many fields and various applications. We will briefly highlight some of the related work in this area as it applies to vegetation detection and/or man-made object detection. We have divided the related work into four groups, based on sensor type.

For image-based detection and classification, one of the most successful approaches has been using the statistics of natural image categories, proposed by Torralba and Oliva.[3] In their work, they create a visual representation and classification rule based on second-order statistics of image categories, scene scale and image scale. Recently, the use of Gaussian Processes (GP) have been used to identify man-made structures within images.[4,5] GP is essentially a generalization of the Gaussian distribution to a process, which is fully specified by its mean and covariance function (instead of vector and matrix, respectively). GP may be used to specify very flexible non-linear regression and have been shown to be successful in image classification problems. Another recent approach, introduced by Kanan and Cottrell,[6] uses a growing trend of biologically-inspired features for image category classification. In their work, they first learn features by applying independent components analysis (ICA) to a dataset of natural color images which produces a set of sparse filters that resemble simple cells in the visual cortex.[7] Then, a saliency map is extracted from an image to create a probability distribution which will be used to inform the selection of feature points for classification. Alvarez et al.[8] propose a method based on convolutional neural networks for road segmentation from a single image. Their segmentation recovers the 3D structure of road scenes that could prove quite useful for obstacle detection and navigation for UGVs. Joulin et al.[9] propose a bottom-up solution to unsupervised image segmentation to divide images into regions consisting of different image classes using discriminative clustering. Domke[10] presents an approach to multi-class image labeling using graphical models to learn conditional random fields achieving state-of-the-art results on the Stanford Background Dataset.[11]

Adding an additional calibrated camera to the perception system allows for stereo-based methods for obstacle detection and navigation. Manduchi et al.[12] use stereo imagery for obstacle detection and segmentation and terrain classification based on a Gaussian mixture modeling approach for autonomous off-road navigation. Rankin et al.[13] describe their approach for terrain classification, pedestrian detection and long range terrain classification using an improved stereo pipeline algorithm. Bajracharya et al.[14] propose their system for building a 3D voxel map based on dense stereo camera range data for obstacle detection and navigation. A team at Southwest Research Institute (SwRI) has demonstrated material classification of a scene, including detecting vegetation, using six spectral cameras as well as texture and statistical data from stereo cameras.[15]

Several methods have been proposed for autonomous vehicle navigation using only 3D laser data. One such approach, introduced by Hebert and Vandapel,[16] uses the 3D points to define local shapes for classification. A continuation of this research, by Vandapel et al.,[17] computes 3D statistics to capture the spatial distribution of points in local neighborhoods for terrain classification. Promising results in each approach are shown, but there are still several issues, such as computational costs and latency, in using 3D laser data for object detection and classification, especially for small vehicles with limited processing capabilities.

To complement the 3D laser data, several researchers have added a camera to the perception system. Dahlkamp et al.[18] describe their approach to fusing data from a laser range finder, pose estimation system and color camera for road detection in desert terrain, which was used in their vehicle for the 2005 DARPA Grand Challenge. Bradley et al.[19] propose a perception system that utilizes several laser scanners and cameras, including a near infrared (NIR) camera to highlight vegetation. Scaramuzza et al.[20] describe an extrinsic self calibration algorithm of a camera and 3D laser range finder, which does not require a predefined calibration pattern or object. A segmentation approach using 3D point cloud data with color information from imagery is proposed by Kim et al.[21] Their approach clusters salient features into meaningful regions with promising results. Strom et al.[22] present a graph-theoretic approach to segmentation for 3D laser data with color information from imagery (or colored 3D laser point clouds). Recently, Markov Random Fields[23] have also been applied to the problem of terrain classification from fused camera and 3D laser range data.

Our research continues in the vein of image-only detection and classification, focused on the problem of vegetation and man-made objects from off-road environments.

(a) (b)

(c) (d)

Figure 1. Example images from a small UGV in an off-road environment

3. TOWARDS VEGETATION AND MAN-MADE OBJECT DETECTION AND CLASSIFICATION

Many algorithms in image segmentation, image classification and multi-class image labeling have been previously proposed. We focus this paper on three methods from the literature. First, we introduce the work of Joulin et al.[9] for image segmentation (or co-segmentation). We then review the work of Kanan and Cottrell[6] on saliency detection using natural scene statistics for image classification. Finally, we introduce and apply the work from Domke[10] to four example images from a small UGV taken in an off-road environment of vegetation and man-made objects. The four images we will use in this paper are shown in Figure 1. The images consist of vegetation, one or more man-made objects (rebar and wooden palate) and shadows.

3.1 IMAGE SEGMENTATION

Research in the area of image segmentation is vast and constantly growing. For our purposes, image segmentation is a means to an end. We wish to segment an image into meaningful regions for classification, so image segmentation may fulfill the detection task of our overall problem. In this section we review a recent method that has achieved state-of-the-art results on several well-known datasets with ground truth for image segmentation. We will then demonstrate the results of this method on our four example images.

Joulin et al.[9] introduce a method for unsupervised image segmentation (or co-segmentation, for multiple images) using discriminative clustering. Co-segmentation involves the process of simultaneously dividing multiple images into k number of classes. For one image, this reduces to the regular segmentation problem. The main idea of using co-segmentation is that using multiple images that contain the same or similar objects may assist with the absence of supervisory information.

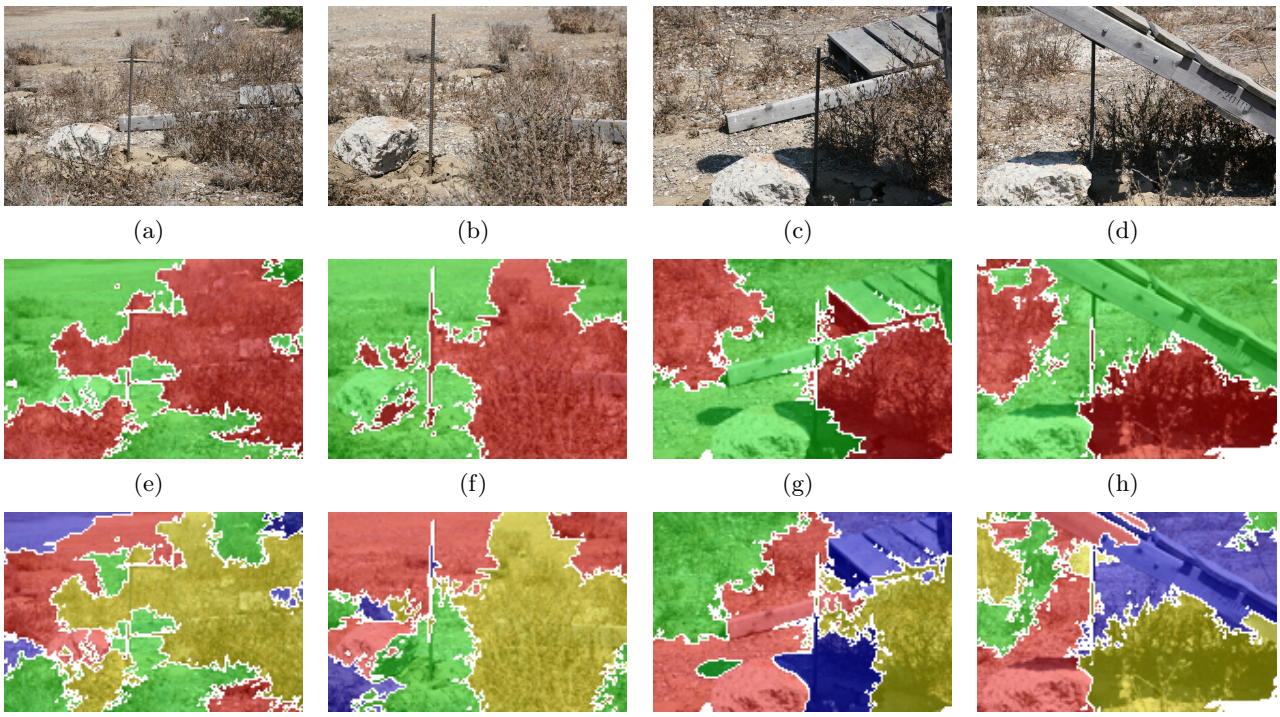


Figure 2. Four example images (a-d) along with their 2-class (e-h) and 4-class (i-l) segmentations

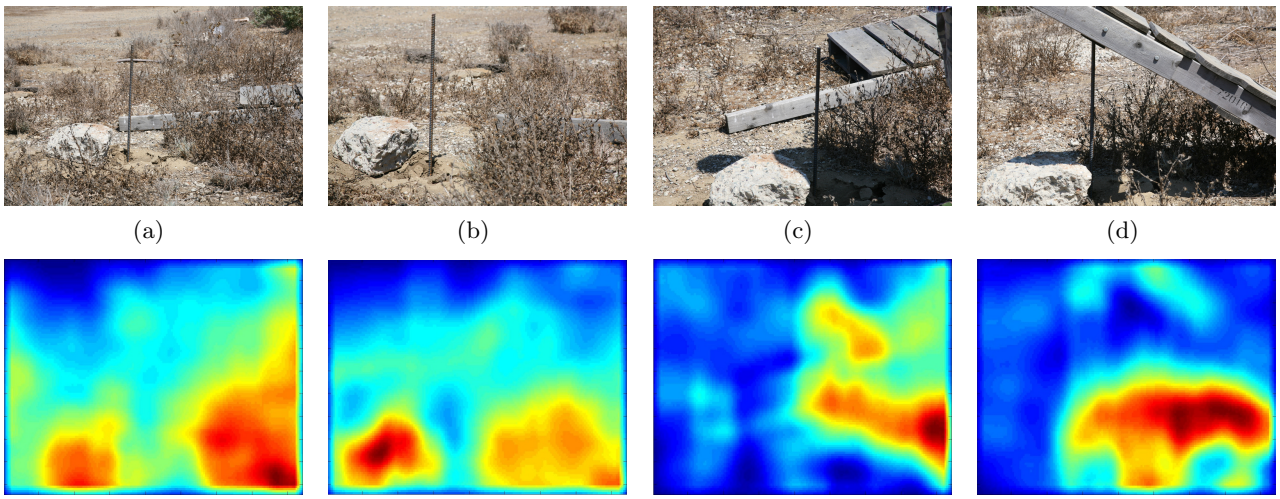
In their framework, discriminative clustering is used for the co-segmentation problem. Discriminative clustering relies on supervised classification (such as the support vector machine (SVM)) techniques to perform unsupervised clustering. Essentially, in a two-class example problem, every pixel in an image is labeled as a foreground or background pixel and discriminative clustering provides a way to learn a classifier given those pixel labels. Then an optimal classifier is found using a least-squares classification framework. The images features used in their work and in ours are scale-invariant feature transform (SIFT)[24] features.

We use the author's provided code for the implementation. Since this method is unsupervised, no additional training is required to perform the segmentation, which is advantageous in situations like ours where training imagery may not be available. The results of their algorithm on our four example images is found in Figure 2. The top row corresponds to the original images. The second and third rows correspond to 2-class and 4-class co-segmentations, respectively. Generally speaking, the algorithm performs quite well at detecting and segmenting vegetation from the image in the 2-class co-segmentation case. In each of those cases, however, the rebar is also grouped together with the vegetation.

3.2 IMAGE SALIENCY

In this section we will briefly introduce the methodology proposed by Kanan and Cottrell[6] for generating image saliency. Saliency, in this context, is simply the regions of an image that are more interesting than others and are based on visual attention. In their work, Kanan and Cottrell generate a saliency map of an image using biologically-inspired features. These features are learned by applying independent components analysis (ICA) to a dataset of natural color images which produces a set of sparse filters that resemble simple cells in the visual cortex.[7] Once the features are learned, they are used in a Saliency Using Natural statistics (SUN) algorithm to generate the bottom-up saliency map. ICA results in features that are largely statistically independent, so a product of unidimensional distributions (modeled as a generalized Gaussian distribution) may be used to model the saliency.

In Kanan and Cottrell's work,[6] they further demonstrate success using a classification framework which uses the saliency map to provide a probability distribution which is used to select feature points for classification.



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 3. Four example images (a-d) along with their saliency maps (e-h)

The performance of their system is comparable to the state-of-the-art in several standard datasets. Since our images have multiple classes in each image, this step is not appropriate.

We applied the SUN algorithm, from the author's provided implementation, to our four example images and the results are shown in Figure 3. The higher the saliency, the closer to red the pixel should appear.

The vegetation in the images is again segmented (or highlighted) from each of the images. For instance, the information in Figures 2e and 3e seems very similar. However, the man-made objects are mostly ignored by the saliency map. One advantage of combining these methods could be seen in Figures 2f and 3f. In Figure 2f, the segmentation algorithm has identified the vegetation region, but has failed to highlight the large stone in our path. In Figure 3f, however, the saliency map has identified the stone as one of the two most important objects in the image. For our uses, the saliency map may be used to inform a segmentation (or co-segmentation) method and/or a classification algorithm.

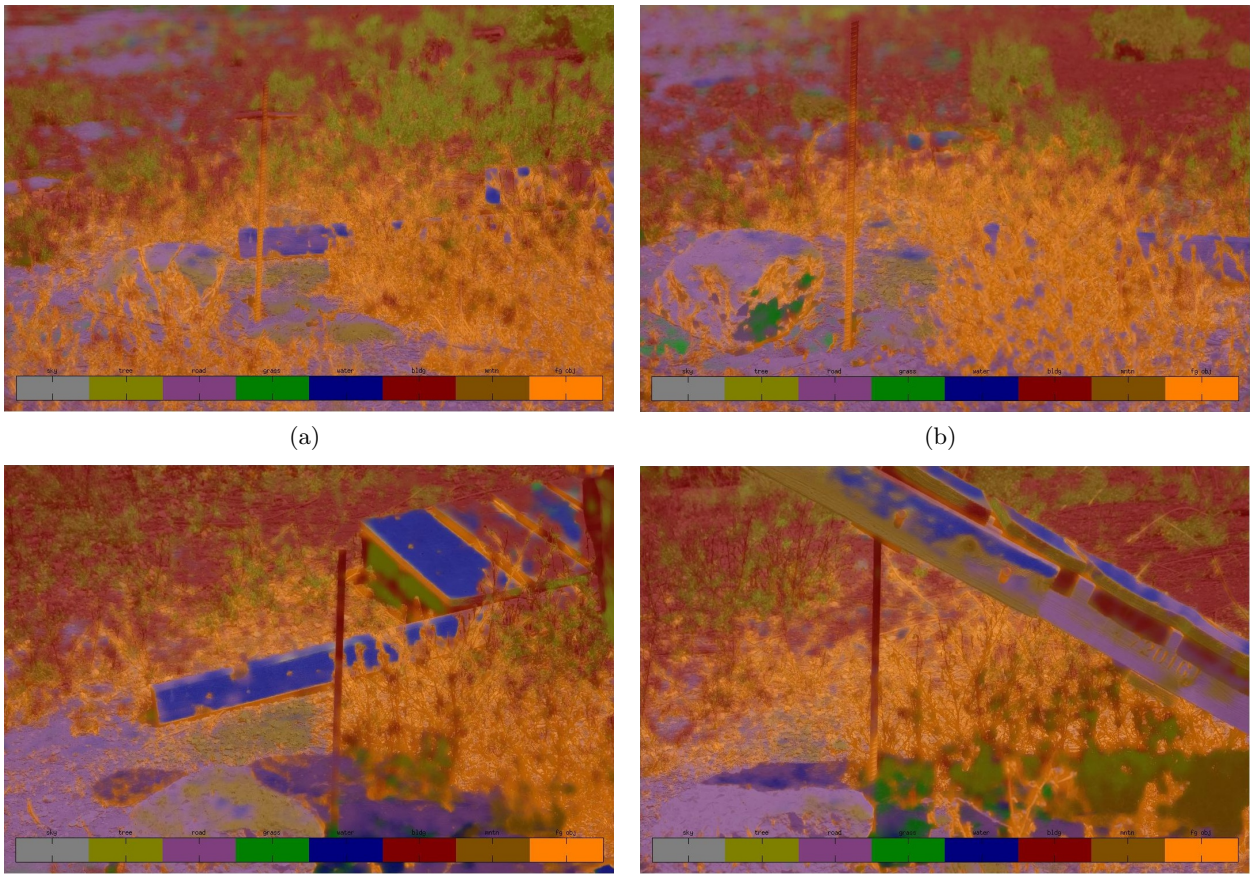
3.3 MULTI-CLASS IMAGE LABELING

In this section we will introduce the work of Domke[10] for multi-class image labeling using graphical models to learn conditional random fields (CRF) for classification.

Domke models an image as a conditional random field (CRF), which has recently become a popular tool for image segmentation and object recognition. A CRF is a variant of a Markov random field, where each random variable may be conditioned on a set of global observations. The parameters of the CRF are usually learned via maximum likelihoods, so most of the previous work has focused approximating the likelihood. In Domke's work, he instead uses "marginalization-based" loss functions to learn the parameters. Two advantages to this approach are that this method is robust to model mis-specification and that the approximation errors are taken into account during the learning process.

Figure 4 displays the results of Domke's method, using the author's Graphical Models Toolbox, on our four example images. In each image, the color corresponds to a class label learned from the Stanford Background dataset.[11] From left to right on the colorbar in the bottom of the image, the classes are sky (gray), tree (olive), road (purple), grass (green), water (blue), building (red), mountain (brown), and foreground object (orange).

As shown in the Figure, Domke's method does quite well at highlighting the foreground objects and even classifies the road pixels correctly, even though the 'road' is not a traditional asphalt road. Additionally, the difficult to detect rebar is classified as foreground and 'building', which we can equate to a man-made object. The vegetation classification has room for improvement. Most of the vegetation in the path of the UGV is classified as foreground, which means that further classification is needed on those pixels. Also, the vegetation



(a) (b)

(c) (d)

Figure 4. Multi-class labels of the four example images, respectively

towards the top (back) of the image has some of the pixels classified as 'building'. Finally, the wooden palate is segmented from the rest of the image, but is classified as either 'road' or 'water'. The majority of these errors may be corrected with a more carefully selected training set, since the Stanford Background dataset is not really indicative of the imagery in our environment. More research is needed to find a general dataset for training models to produce acceptable results on imagery such as the examples provided.

4. CONCLUSION AND FUTURE WORK

The purpose of this work is to highlight recent advances in computer vision towards the goal of vegetation and man-made object detection and classification. We believe that the most relevant advances are in three areas: image segmentation, saliency and multi-class image labeling. We have presented a state-of the art method in each area and shown the results of that method applied to four example images taken from a small UGV in an off-road environment. There are several promising qualitative results, such as vegetation being segmented and/or highlighted in an image in the first two works described. However, detecting man-made objects consistently was not achieved. Therefore, there are several future directions for this research.

First, we believe there are opportunities to fuse the three methods, and possibly others, to form a more robust approach. For instance, the salient regions in the second method only overlap with some of the vegetation from the first, so they give us slightly different, and possibly complementary, information. Second, since we are dealing with images that are not likely to coincide with those found in abundance on the internet, training for multi-class image labeling becomes difficult. Avenues such as semi-supervised classification might be worth pursuing so that an expert is involved in some of the labeling. Third, none of the algorithms tested, or those researched, may

be run in real-time. More investigation is needed into reducing computation requirements. Another possible direction for this research is to fuse color imagery with an infrared or near infrared sensor,[25,26] which is also passive, light-weight and low power.

REFERENCES

[1] Thrun, S., Montemerlo, M., Dahlkamp, H., Stavens, D., Aron, A., Diebel, J., Fong, P., Gale, J., Halpenny, M., Hoffmann, G., et al., "Stanley: The robot that won the darpa grand challenge," *Journal of field Robotics* **23**(9), 661–692 (2006).

[2] Thrun, S., "Toward robotic cars," *Communications of the ACM* **53**(4), 99–106 (2010).

[3] Torralba, A. and Oliva, A., "Statistics of natural image categories," *Network: computation in neural systems* **14**(3), 391–412 (2003).

[4] Zhou, H. and Suter, D., "Fast sparse gaussian processes learning for man-made structure classification," in [*IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*], 1–6, IEEE (2007).

[5] Rasmussen, C. E., "Gaussian processes for machine learning," (2006).

[6] Kanan, C. and Cottrell, G., "Robust classification of objects, faces, and flowers using natural image statistics," in [*2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*], 2472–2479, IEEE (2010).

[7] Caywood, M. S., Willmore, B., and Tolhurst, D. J., "Independent components of color natural scenes resemble v1 neurons in their spatial and color tuning," *Journal of Neurophysiology* **91**(6), 2859–2873 (2004).

[8] Alvarez, J. M., Gevers, T., LeCun, Y., and Lopez, A. M., "Road scene segmentation from a single image," in [*Computer Vision–ECCV 2012*], 376–389 (2012).

[9] Joulin, A., Bach, F., and Ponce, J., "Discriminative clustering for image co-segmentation," in [*2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*], 1943–1950, IEEE (2010).

[10] Domke, J., "Learning graphical model parameters with approximate marginal inference," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **PP**(99), 1–1 (2013).

[11] Gould, S., Fulton, R., and Koller, D., "Decomposing a scene into geometric and semantically consistent regions," in [*2009 IEEE 12th International Conference on Computer Vision*], 1–8, IEEE (2009).

[12] Manduchi, R., Castano, A., Talukder, A., and Matthies, L., "Obstacle detection and terrain classification for autonomous off-road navigation," *Autonomous robots* **18**(1), 81–102 (2005).

[13] Rankin, A., Bajracharya, M., Huertas, A., Howard, A., Moghaddam, B., Brennan, S., Ansar, A., Tang, B., Turmon, M., and Matthies, L., "Stereo-vision-based perception capabilities developed during the robotics collaborative technology alliances program," in [*SPIE Defense, Security, and Sensing*], 76920C–76920C, International Society for Optics and Photonics (2010).

[14] Bajracharya, M., Ma, J., Howard, A., and Matthies, L., "Real-time 3d stereo mapping in complex dynamic environments," in [*International Conference on Robotics and Automation - Semantic Mapping, Perception, and Exploration (SPME) Workshop*], (2012).

[15] Lamm, R., "Unmanned and downrange." `http://www.swri.org/3pubs/ttoday/Summer12/pdfs/UnmannedandDownrange.pdf` (Summer 2012).

[16] Hebert, M. and Nicolas, V., "Terrain classification techniques from ladar data for autonomous navigation," in [*In Collaborative Technology Alliances Conference*], (2003).

[17] Vandapel, N., Huber, D. F., Kapuria, A., and Hebert, M., "Natural terrain classification using 3-d ladar data," in [*IEEE International Conference on Robotics and Automation (ICRA)*], **5**, 5117–5122, IEEE (2004).

[18] Dahlkamp, H., Kaehler, A., Stavens, D., Thrun, S., and Bradski, G., "Self-supervised monocular road detection in desert terrain," in [*Proc. of Robotics: Science and Systems (RSS)*], (2006).

[19] Bradley, D. M., Unnikrishnan, R., and Bagnell, J., "Vegetation detection for driving in complex environments," in [*IEEE International Conference on Robotics and Automation*], 503–508, IEEE (2007).

[20] Scaramuzza, D., Harati, A., and Siegwart, R., "Extrinsic self calibration of a camera and a 3d laser range finder from natural scenes," in [*IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*], 4164–4169, IEEE (2007).

[21] Kim, G., Huber, D., and Hebert, M., "Segmentation of salient regions in outdoor scenes using imagery and 3-d data," in [*IEEE Workshop on Applications of Computer Vision (WACV)*], 1–8, IEEE (2008).

[22] Strom, J., Richardson, A., and Olson, E., "Graph-based segmentation for colored 3d laser point clouds," in [*IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*], 2131–2136, IEEE (2010).

[23] Häselich, M., Lang, D., Arends, M., and Paulus, D., "Terrain classification with markov random fields on fused camera and 3d laser range data," in [*Proceedings of the 5th European Conference on Mobile Robotics (ECMR)*], 153–158 (2011).

[24] Lowe, D. G., "Object recognition from local scale-invariant features," in [*The proceedings of the seventh IEEE international conference on Computer vision, 1999*], **2**, 1150–1157, Ieee (1999).

[25] Varjo, S., Hannuksela, J., and Alenius, S., "Comparison of near infrared and visible image fusion methods.," in [*Proc. International Workshop on Applications, Systems and Services for Camera Phone Sensing*], (2011).

[26] Krotosky, S. and Trivedi, M., "Registration of multimodal stereo images using disparity voting from correspondence windows," in [*IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*], 91–91, IEEE (2006).