

# On-Line Learning of Long-Range Obstacle Detection for Off-Road Robots

Raia Hadsell<sup>1</sup>, Pierre Sermanet<sup>2</sup>, Jan Ben<sup>2</sup>, Jeff Han<sup>1</sup>, Sumit Chopra<sup>1</sup>,  
Marc’Aurelio Ranzato<sup>1</sup>, Yury Sulsky<sup>1,2</sup>, Beat Flepp<sup>2</sup>, Urs Muller<sup>2</sup>, Yann LeCun<sup>1</sup>

(1) Courant Institute of Mathematical Sciences, New York University

(2) Net-Scale Technologies, Morganville, NJ 07751, USA

The method of choice for vision-based driving in off-road mobile robots is to construct a traversability map of the environment using stereo vision. In the most common approach, a stereo matching algorithm, applied to images from a pair of stereo cameras, produces a “point-cloud”, in which the most visible pixels are given an XYZ position relative to the robot. A traversability map can then be derived using various heuristics, such as counting the number of points that are above the ground plane in a given map cell. Maps from multiple frames are assembled in a global map in which path finding algorithms are run [2, 3, 1]. The performance of such stereo-based methods is limited, because stereo-based distance estimation is often unreliable above 8 or 10 meters (for typical camera configurations and resolutions). This may cause the system drive as if in a self-imposed “fog”, driving into dead-ends, and taking time to discover distant pathways that are obvious to a human observer.

We present two different solutions to the problem of long-range obstacle detection and path planning. Experiments were run on the LAGR robot platform, setting the cameras resolution at 320x240.

**Method 1: Computing Polar Traversability Map from Stereo.** The practical range of simple stereo-based map building is limited for two reasons: (1) it is difficult to estimate whether far-away points are near the ground or above the ground; (2) the distance estimates are quite inaccurate for points more than 7 or 8 meters away from the camera. To solve problem 1, we estimate the parameters of the ground plane by fitting a plane through the stereo point cloud. Two methods were used: Hough transform on point clouds in elevation, azimuth, disparity space; and EM-like robust plane fitting on point clouds in XYZ space. The traversability of an area is estimated by measuring the density of points that are above the ground plane in that area. Problem 2 is approached by noting that, while absolute range estimates of distant points are inaccurate, relative range estimates are relatively accurate, and azimuth estimates are very accurate. This suggests that searching for good direction in which to drive the robot is better performed using a map of the visible environment represented in *polar coordinates*, rather than using a cartesian map of the entire ground. Our system identifies candidate waypoints up to 15 meters away in this local polar map, and uses them as starting points of a path finding algorithm in a global cartesian map. The path finding algorithm is a new approximate A\*-like method based on ray casting, dubbed “raystar”. The system drives significantly faster than the LAGR baseline system on various test runs (videos will be shown).

**Method 2: On-Line Learning from Distance-Normalized Monocular Images.** Humans can easily locate pathways from monocular views, e.g. trails in a forest, holes in a row of bushes. Method 2 is an attempt to use on-line learning to provide the same capability to a mobile robot. Although supervised learning can be used for robot driving [4], autonomous learning is far preferable. One long-advocated idea for autonomous learning in robots is to use the output of reliable modules (such as traversability from stereo at short range) to provide labels for a trainable module (such as a long-range obstacle detector). In one spectacular demonstration of this idea, short-range traversability data was used to train a mixture of Gaussians model of the RGB color of traversable areas on-the-fly [5].

Our proposed approach, designed for the LAGR robot, builds a *distance-invariant pyramid of images at multiple scales*, such that the appearance in the image at scale  $X$  of an object sitting on the ground  $X$  meters away is identical to the appearance in the image of scale  $Y$  of the same object when sitting on the ground  $Y$  meters away.

First, stereo data is used to label the traversability of visible areas up to 10 meters. Then, the labels are used to train a discriminative classifier *at every frame* which maps sub-windows in the image pyramid to the computed traversability at the corresponding location on the ground. The classifier is then applied to images in the pyramid from 10 meters to 30 meters (far beyond stereo range).

To build the image pyramid, differently sized sub-images are cropped from the original RGB frame such that each is centered around a specific (imaginary) line on the ground that is a given distance from the robot’s camera. Each extracted sub-image is then subsampled to make it a uniform height (12 pixels), resulting in image bands in which the appearance of an object on the ground is independent of its distance from the camera (only the band in which it appears varies, see figure 1). These uniform-height, variable-width bands form a size-normalized image pyramid whose 20 scales are separated by a factor of  $2^{\frac{1}{4}}$ . In order to extract horizon-leveled sub-images from exact ranges, a plane must be fit to the image first. This is done using a Hough transform.

By classifying windows taken from the bands of the image pyramid, traversability information can be directly mapped to specific world coordinates, since the distance to the center of each band is known. Thus, the pyramid provides the structure so that the long-range obstacle detector (OD) can generate accurate range maps.

The long-range OD goes through a labeling, training, and classification cycle on every frame. First, each overlapping 12x3 pixel RGB window from the right camera is assigned a traversability label (ground or obstacle) if it is within stereo

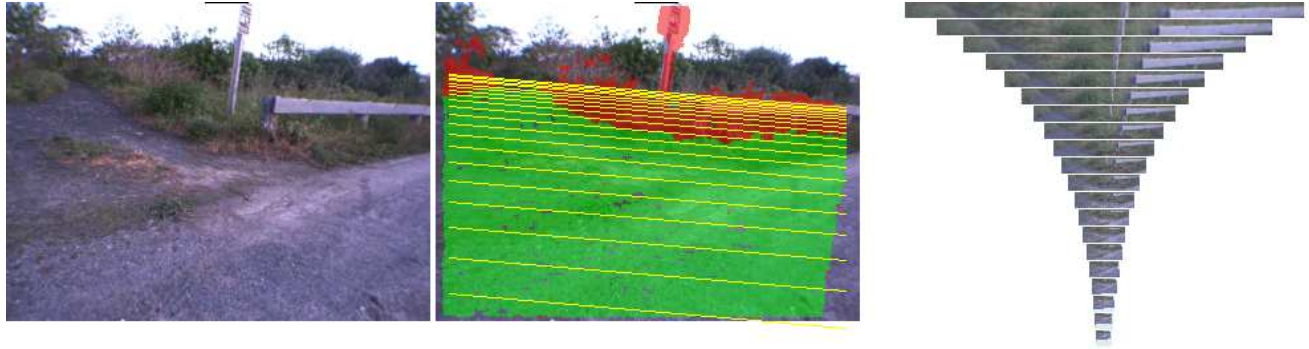


Figure 1: *left* The input image from one camera. *center* The input image with stereo labels overlaid in green and red. Horizon-leveled focus-lines are shown in yellow. *right* The size-normalized image pyramid. The number to the left of each band corresponds to the distance (in meters) from the robot to the focus line that is at the center of that band.

range ( $< 10m$ ) and if stereo data is available. Then feature vectors are computed for all windows over the entire pyramid. Each feature vector is comprised of euclidean distances or correlations between a  $12 \times 3$  RGB window and 100 fixed prototypes trained in advance with an unsupervised learning algorithm. A logistic regression classifier is applied to the feature vectors, and trained using the labels provided by stereo. The resulting classifier is then applied to all feature vectors in the pyramid, including those with stereo labels.

Figure 2 shows examples of the maps generated by the long-range obstacle detector. The long-range OD not only yields surprisingly accurate traversability information at distance up to 30 meters (far beyond stereo range), but also produces smooth, dense traversability maps for areas that are within stereo range. The stereo-derived maps often have noisy spots or holes - disastrous to a path planner - but the adaptive long-range OD produces maps that are smooth and accurate, without holes or noise.

Videos of the robot traversing various courses will be shown. The behavior of the LAGR baseline system, the proposed stereo-based system, and the proposed long-range OD system will be discussed. The proposed stereo-based is typically 2 to 4 times faster than the baseline system. Quantitative data on the long-range OD system will be presented at the workshop.

## References

- [1] S. B. Goldberg, M. Maimone, and L. Matthies. Stereo vision and robot navigation software for planetary exploration. *IEEE Aerospace Conf. Proc.*, March 2002. 1
- [2] A. Kelly and A. Stentz. Stereo vision enhancements for low-cost outdoor autonomous vehicles. *Int'l Conf. on Robotics and Automation, Workshop WS-7, Navigation of Outdoor Autonomous Vehicles*, May 1998. 1
- [3] D. J. Kriegman, E. Triendl, and T. O. Binford. Stereo vision and navigation in buildings for mobile robots. *IEEE Trans. Robotics and Automation*, 5(6):792–803, 1989. 1
- [4] Y. LeCun, U. Muller, J. Ben, E. Cosatto, and B. Flepp. Off-road obstacle avoidance through end-to-end learning. In *Advances in Neural Information Processing Systems (NIPS 2005)*. MIT Press, 2005. 1
- [5] D. Leib, A. Lookingbill, and S. Thrun. Adaptive road following using self-supervised learning and reverse optical flow. *Proceedings of Robotics: Science and Systems*, June 2005. 1

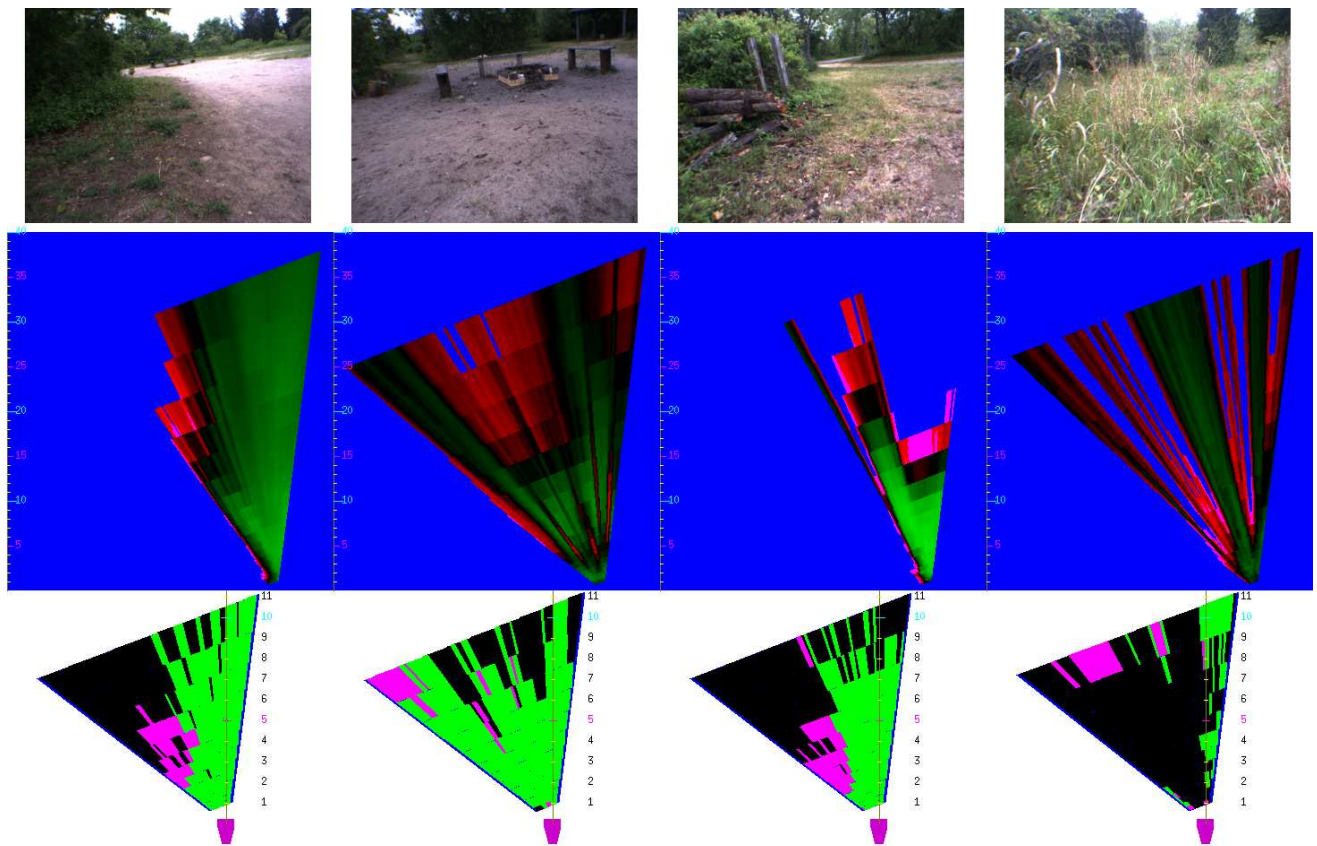


Figure 2: Examples of performance of the long-range obstacle detector. Each example shows the input image (top), the map produced by the long-range OD (middle), and the stereo labels (lower). Note that the stereo (lower map) is shown at a range of 0-11m, whereas the long-range OD map is shown from 0-40m.