2008

# LADAR-based Pedestrian Detection and Tracking

Luis E. Navarro-Serment
*Carnegie Mellon University*

Christoph Mertz
*Carnegie Mellon University*

Nicolas Vandapel
*Carnegie Mellon University*

Martial Hebert
*Carnegie Mellon University*



Published In

Proc. 1st. Workshop on Human Detection from Mobile Robot Platforms, IEEE ICRA 2008.

# LADAR-based Pedestrian Detection and Tracking

Luis E. Navarro-Serment, Christoph Mertz, Nicolas Vandapel, and Martial Hebert, *Member, IEEE*

***Abstract*—The approach investigated in this work employs LADAR measurements to detect and track pedestrians over time. The algorithm can process range measurements from both line and 3D scanners. The use of line scanners allows detection and tracking at rates up to 75 Hz. However, this type of sensor may not always perform satisfactorily in uneven terrains. A 3D LADAR is used to improve operation in uneven terrains, by first estimating the local ground elevation, and then performing the detection using the measurements corresponding to a certain height above the ground. The information pipeline used to feed sensor data into the algorithm is the same for both types of sensors. The perceptual capabilities described aim to form the basis for safe and robust navigation in robotic vehicles, necessary to safeguard pedestrians operating in the vicinity of a moving robotic vehicle.**

## I. Introduction

THE ability to avoid colliding with other moving objects is particularly important in autonomous vehicles. This is especially important in cases where the vehicle operates in close proximity with people. In order to be effective, a vehicle's collision avoidance system must perform two basic tasks: detection and tracking of moving objects. The timely detection of an object makes the vehicle aware of a potential danger in its vicinity. Similarly, the vehicle can predict the most likely near future positions of an object being tracked, and make corrections to its present course accordingly.

Robust and reliable detection and tracking has attracted a lot of attention in recent years, driven by applications such as pedestrian protection [1], vehicle platooning, and autonomous driving [2]. This is a difficult problem, which becomes even harder when the sensors (e.g., optical sensors, radar, laser scanners) are mounted on the vehicle rather than being fixed, such as in traffic monitoring systems. Effective detection and tracking require accurate measurements of object position and motion, even when the sensor itself is moving. Range sensors are well suited to this problem because a first-order motion correction can be made by simply subtracting out self-motion from range measurements. Unfortunately, merely subtracting out ego-motion does not eliminate all the effects of motion because the perceived object's shape changes as different aspects of the object come into view, and this change can easily be misinterpreted as motion. Moreover, the perceived appearance of an object depends on its pose, and can also be affected by nearby objects. Finally, complex outdoor environments frequently involve cluttered backgrounds, susceptibility to ground returns in uneven terrains, and unpredictable interaction between traffic participants.

There is a large body of work done using laser line scanners as the primary sensor for pedestrian detection and tracking. In our group, we have developed systems using SICK™ laser line scanners to collect range measurements. These implementations produced satisfactory results [3]. The line scanners generate data at fast rates, which is also easily processed. However, as shown in Fig. 1, the data collected may contain ground returns, which increase the probability of false detections. As seen in the figure, the line scanner mounted on the left side of the vehicle has a sensing plane that intersects the ground in uneven terrain. This sensor has no way of knowing whether a measurement corresponds to the ground or to a pedestrian. In contrast, a 3D LADAR (i.e. one who produces a set of 3D points, or point cloud) can produce a more complete representation of the environment, from which both object and terrain information can be extracted. As shown in Fig. 1, the 3D LADAR in front of the vehicle can collect measurements from both the human and his surrounding environment. One way of exploiting this information is by generating a ground elevation map as the vehicle traverses the environment, so that ground returns can be eliminated, and the pedestrians are more robustly identified.

In this paper, we describe a strategy to detect and track moving objects in real-time. The approach detects humans

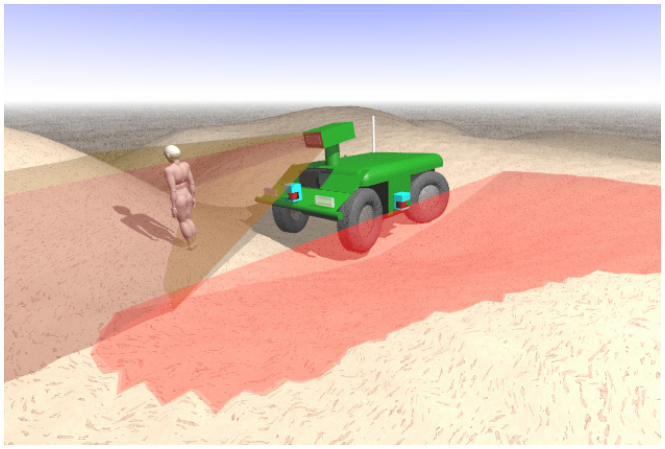

Fig. 1. Overview of the detection and tracking system. At the center of the figure is the vehicle, which carries multiple line scanners (one on each side, as shown here). A 3D LADAR, mounted on top of the vehicle, can scan the environment in multiple directions using a pan-tilt platform.

This work was conducted through collaborative participation in the Robotics Consortium sponsored by the U. S. Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-209912.

The authors are with the Robotics Institute, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213 USA (lenscmu@ri.cmu.edu, cmertz@andrew.cmu.edu, vandapel@cs.cmu.edu, hebert@ri.cmu.edu ).

from LADAR measurements and tracks them over time. The tracker detects moving objects and estimates their position and motion, while largely ignoring self-motion-induced changes in the range measurements. In particular, we focus on the description of the incorporation and use of 3D LADAR measurements.

Our work differs from previous approaches in that our system is capable of processing range measurements from both line and 3D scanners. The information pipeline used to feed sensor data into our system is the same for both types of sensors, which simplifies fusing data from multiple scanners. The use of 3D LADAR improves the performance in uneven terrains, while the line scanners, which usually operate at faster rates, perform better in flatter areas [3]. This feature allows the system to adapt to a wider range of operation environments by combining the advantages of each sensor type. We present experimental results of detection performance using 3D LADAR. These results were obtained from field tests performed on a Demo III XUV [4] , which is the target platform (Fig. 2).

## II. Related Work

The most commonly used approaches to detection and tracking of moving objects for vehicular applications involve both passive and active range sensors [5]. Passive sensors, such as video cameras, acquire data in a non-intrusive way. In [6] the authors present an extensive review of vision-based on-road vehicle detection systems.

Active range sensors, such as radar and LADAR, are capable of measuring distances directly without requiring powerful computing resources. In particular, recent models of laser scanners are capable of gathering high resolution data at high scanning speeds, and are available in enclosures suitable for vehicular applications. The closest work related to our approach involves the use of laser line scanners. In [7], the authors describe the application of a multilayered laser scanner for pedestrian classification. Vehicle odometry is used to estimate self-motion, to remove the effects of sensor motion. A Kalman filter estimates the object velocity. Tracked objects are classified as car, pedestrian, etc., based on their shape and behavior over time. This group also produced a second system [8], in which a Kalman filter estimates motion based on the change in position of an object's estimated center-point. Object classification is used to fit a class-specific prior rectangular model to the points. Although not mentioned explicitly, this appears to be an approach to reducing shape-change motion artifacts. The success of this technique depends on the correctness of the classification. Each object class also has distinct fixed Kalman filter parameters. A multi-hypothesis approach is used to mitigate the effect of classification error. The emphasis of both efforts is on single-LADAR systems, and multi-scanner fusion is not considered.

Fig. 2. The Demo III Experimental Unmanned Vehicle (named XUV) used in our tests. The vehicle shown is equipped with 4 line scanners (one on each side), and a 3D LADAR mounted on a pan-tilt platform.

Using an occupancy grid appears to be a convenient way of detecting moving objects by simply observing the changes in occupancy values for each location. However, maintaining an occupancy grid is expensive; [9] addressed this problem with a sparse representation of open space. Yet, one cannot disregard the possibility that an object was there already but could not be detected due to occlusion or because it was out of range.

In [10], motion is measured by registering old and new scans using chamfer fitting. A constant-velocity, constant-angular velocity Kalman filter is used. Because the scanner is placed above the leg level, a rigid body model is satisfactory. Although this paper does not use moving scanners, it is noteworthy because of its attempt to quantitatively evaluate performance without ground truth by measuring the position noise of stationary tracks, the measurement residue of moving tracks, and the occurrence of false positive and false negative errors in moving object detection.

A system that deals particularly well with crowds of people is described in [11]. The authors use a feature extraction method based on accumulated distribution of successive laser frames. The final tracker is based on the combination of independent Kalman filter and Rao-Blackwellized Monte Carlo data association filter.

Some systems use 3D laser data for tracking. In [12] 3D scans are automatically clustered into objects and modeled using a surface density function. A Bhattacharya similarity measure is optimized to register subsequent views of each object enabling good discrimination and tracking, and hence detection of moving objects.

Mapping is a related field, especially if the mapping is done in dynamic environments [13]. E.g. in [14] the authors generalize Simultaneous Localization and Mapping (SLAM) to allow detection of moving objects, relying primarily on the scanner itself to measure self-motion. An extended Kalman filter with a single constant velocity model is used in a multi-hypothesis tracker. As opposed to our work, their emphasis appears to be on mapping in the presence of moving objects, rather than the real-time detection of moving objects when no map is needed.

## III. Algorithm Description

In this section, the algorithm for pedestrian detection and tracking is described. A more detailed description, particularly of the elements used only with true line scanners for the generation of collision warnings, is presented in [16].

Objects in the vicinity of the vehicle are detected using measurements collected by the vehicle's LADAR scanners, as shown in Fig. 1. As seen in the figure, the line and 3D scanners differ in their respective sensing areas. We have designed our algorithm to accommodate both types of sensors through the use of the *scan line* as the basic data element. This is simply a vector of distance measurements coming from consecutive bearings, which is exactly the kind of data provided directly by a line scanner such as the SICK™ laser scanner.

### A. Segmentation, Tracking, and Human Detection

The algorithm processes the scan line in the same way, independently of which type of sensor produced it. As shown in Fig. 3, a scan line enters the common processing steps of segmentation, tracking, and human detection. The difference resides in the way the scan line is obtained: a line scanner produces an *actual scan line* directly, while a *virtual scan line* is extracted from the point cloud generated by a 3D LADAR. The virtual scan line imitates an actual scan line collected from a line scanner. This is done by projecting the 3D measurements into a 2D plane. However, in order to take advantage of the more abundant information contained in the point cloud, the projection is done after computing an estimation of the ground elevation, and removing the measurements corresponding to ground returns (this process is described in a subsequent section).

As seen in Fig. 3, once a scan line is obtained (either directly or through a projection in the 2D plane), it is processed through the sequence: object segmentation, object tracking, and human detection. This is done with each scan line. An illustrative example using experimental data showing these steps is presented in Fig. 4.

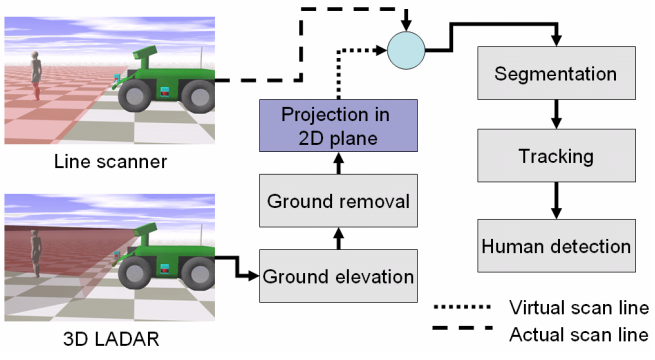


Fig. 3. Overview of the detection and tracking process. A scan line is used as the basic data element, which undergoes the common processes of segmentation, tracking, and human detection.

#### 1) Object Segmentation

The first step is segmentation, which is the grouping of the points contained in the scan line (i.e. range and bearing measurements) into potential objects. The scan line contains a sequence of points (Fig. 4-(a)), and the segmentation process determines which points are likely to belong to the same object, as shown in Fig. 4-(b). We consider that two neighboring points belong to the same object if their separation is less than 0.8 m.

Depending on its size and shape, each object can be summarized by different features, as described in [16]. In this application, given its focus on pedestrians, each object is represented by its center feature, which is the center of the bounding box enclosing the points belonging to the object, as shown in Fig. 4-(c).

#### 2) Object Tracking

The objects detected in the current scan are then tested for association with objects from previous scans. If the bounding boxes of two objects from current and past scans overlap, they are considered to be the same object. For two objects to overlap, at least one actual measurement point must fall inside the object's bounding box, taking into account its predicted motion path. The motion of the center feature (calculated from the match) is then fed into a Kalman filter, which estimates the velocity of the object. As shown in Fig. 4-(d), the object enclosed by the ellipse was matched with the same object detected in a previous scan, using the motion predicted from where the object was before, as seen in Fig. 4-(c).

At each iteration cycle, a sequence of validation tests is performed to determine if a candidate object is a valid target. For example, an estimate has to undergo a minimum number of cycles; the standard deviation of the estimate should not exceed a given maximum threshold, and the estimate should remain consistent for at least a given minimum time. These tests increase robustness against noisy measurements.

To track objects using multiple sensors, the algorithm is executed over all the participating sensors arranged around the vehicle with overlapping fields of view. As a result, it is necessary to "hand off" objects tracked in one field of view to the next. The fusion of sensors happens at the object level. There is only one object list and each scan updates the objects within its own field of view.

#### 3) Human Detection

All objects currently being tracked are tested to determine whether they are humans or not. For this purpose, each target is evaluated by computing its *Strength of Detection* (SOD). The SOD is a measure of how confident the algorithm is that the detected object is actually a human. The objects evaluated with a SOD larger than a given threshold are classified as Humans.

The use of the SOD is motivated by several goals: first, the SOD supports the fusion of detections from different

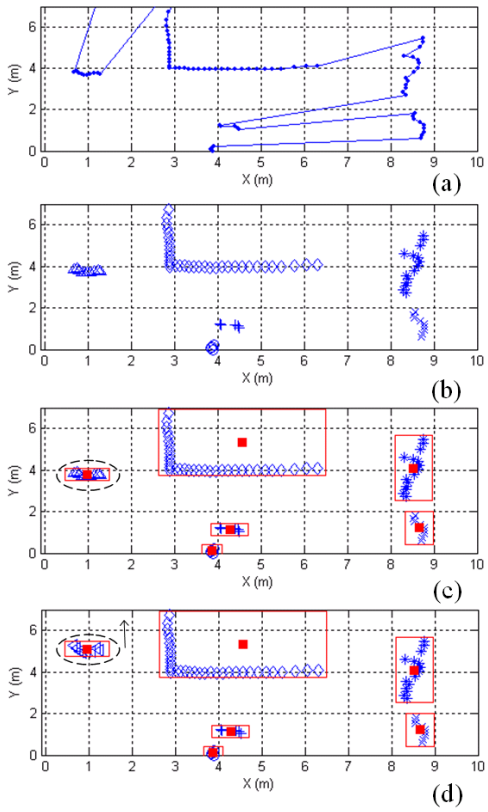


Fig. 4. A sensor placed at (0, 0) produces a scan line (a). The segmentation step groups points into potential objects, as indicated by the sets of points with similar markers (b). The objects are summarized by the center of their bounding boxes (c), whose motion is fed into a Kalman filter for tracking (d).

systems by serving as a common currency for comparison. Second, different applications may need to increase or relax the rigorousness of the detection uncertainty to adapt to the mission requirements. For example, a vehicle moving among the crowds in a shopping center may apply a lower threshold on the SOD to ensure that no pedestrians get injured, while a vehicle operating off-road in a rural area will navigate less hesitantly by applying a higher threshold on the SOD. Third, the comprehensive nature of the SOD helps to maintain a low count in the number of parameters to adjust when tuning the algorithm for other sensors.

*a) Calculation of the SOD*

There are four measures that go into the SOD: The size, the distance traveled ($d_t$), the variation in the size ($\sigma_s$), and the velocity ($\sigma_v$). The variation is the variance computed over the last 14 cycles. The size test discriminates it from large objects like cars and walls. The distance traveled test discriminates against stationary objects like barrels and posts. The variation tests discriminate against vegetation, since their appearance changes a lot due to their porous and flexible nature.

The SOD is the product of the scores of the size and distance traveled test, and the square root of the scores of the variance tests:

$$SOD = S_{d_t} S_{size} \sqrt{S_{\sigma_v} S_{\sigma_s}} \quad (1)$$

We take the square root of the variance tests so that together they have the same weight as the other two tests.

For size and variance tests the score is calculated as follows:

$$S_{size,\sigma_s,\sigma_v}(x) = \begin{cases} 0 & x \geq a_e \\ (a_e - x)/(a_e - a_b) & a_e > x > a_b \\ 1 & x \leq a_e \end{cases} \quad (2)$$

and for the distance traveled test the score is

$$S_{d_t}(x) = \begin{cases} 0.75\dfrac{a_b + x}{2a_b} & S_1 < 1, x < a_b \\ 0.75 & S_1 < 1, x \geq a_b \\ 0.75 + \dfrac{x}{4a_m} & S_1 = 1, a_m > x > a_b \\ 1 & \text{else} \end{cases} \quad (3)$$

where

$$S_1 = S_{size} \sqrt{S_{\sigma_s} S_{\sigma_v}} \quad (4)$$

The values used for the thresholds $a$ are listed in Table I.

TABLE I
THRESHOLDS FOR SOD TESTS

| | $a_b$ | $a_e$ | $a_m$ |
|---|---|---|---|
| Size [m] | 1.0 | 2.0 | - |
| Size variation [m$^2$] | 0.035 | 0.45 | - |
| Velocity variation [(m/s)$^2$] | 0.01 | 0.1 | - |
| Distance traveled [m] | 1.5 | - | 3.0 |

The reasoning behind these calculations is as follows. The $a_b$ values are the thresholds for the best accuracy, i.e. the values that maximize the weighted average of true positive and true negative rates. $a_e$ are the extreme values, i.e. the largest values we observed for humans. For the size the numbers agree with straightforward intuition: 1 m is slightly larger than a normal cross section of a person and 2 m is the span of outstretched arms.

The calculation for distance traveled test is different. In contrast to the others, the larger the distance traveled is, the better. But also, a person can have a distance traveled of 0 m, so there is no upper or lower limit that can discriminate against a person; instead, it is possible to discriminate against fixed objects. $a_m$ is the extreme value we observed for fixed objects. Those objects can have the appearance of movement because of miss-association.

## B. Projection in 2D Plane

As shown in Fig. 3, a 3D scanner produces a point cloud, from which a "slice" is projected onto the 2-D plane,

resulting in a virtual scan line. This is done by collapsing onto the plane all the points residing within the slice, which is defined by its height above the ground. This projection imitates the scan line produced by a SICK™ laser line scanner, and conveniently allows the system to read and process this measurement in the same way as the other sensors. The concept is illustrated in Fig. 5.

*1) Elevation Map*

The system accumulates processed LADAR measurements in a grid, which is always aligned with the vehicle's reference frame, and is centered on the vehicle's current position. The elevation map provides a surface description model in which each cell contains an estimate of the ground elevation for the corresponding area. The elevation is computed by averaging the height of all scan points that are inside the cell. To keep the computational cost of maintaining the map to a minimum, the map is indexed as a 2D ring buffer using modulo arithmetic. This logically scrolls the map as the vehicle moves by physically wrapping around in memory [15]. In this way, only the cells for which no estimates have been computed are updated when the map is scrolled in response to vehicle movement.

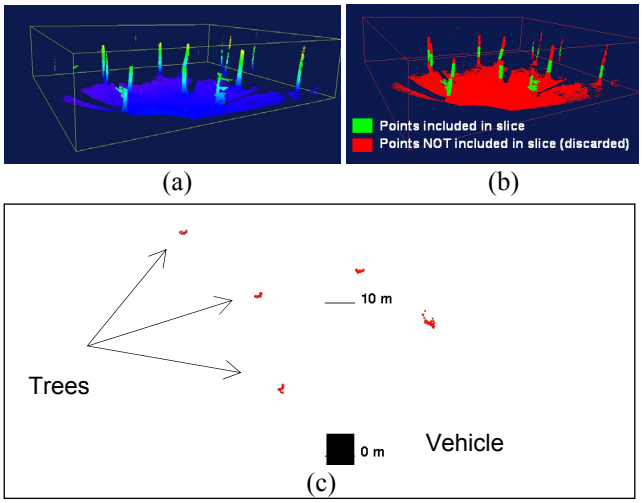


Fig. 5. Projection of virtual scan line. A point cloud is collected from the environment depicted in Fig. 2 (a). The points in the cloud located within a certain height from the ground are projected into a 2D plane (b), and processed as if it were a single scan line. The resulting projection is shown in (c), top view. As expected, this measurement resembles what would be obtained using a line scanner.

*2) Ground Removal*

The abundance of environmental information contained in the point cloud is exploited by generating a ground elevation map as the robot traverses the environment. The system is able to adapt to different environments by varying the shape of the sensing plane, i.e., by adjusting the height of the slice from which points are projected onto a two-dimensional plane. We use the ground elevation to search for potential targets at a constant height above the ground, while ignoring spurious measurements produced by ground returns. Ground returns are greatly reduced, increasing the likelihood of collecting measurements from potential targets, while reducing the probability of false detections resulting from processing ground returns (Fig. 6).

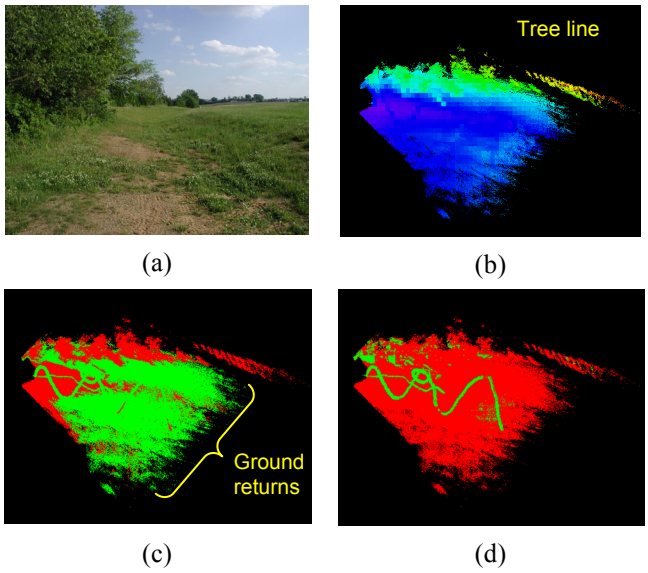


Fig. 6. Ground removal. This experiment was conducted in uneven terrain (a). The corresponding ground elevation map is shown in (b) (viewed from above), where hot colors indicate the highest elevations. The projected 2D slices are shown in figures (c) and (d). Points used in the slice are shown in green, while points in red are ignored. In (c), the virtual slice is just a plane, and the lack of compensation for ground elevation results in significant ground returns, as evidenced by the large number of points included in the slice. Conversely, in (d) the virtual slice is adjusted depending on ground elevation. The trajectory of a pedestrian is now easily identified.

## IV. Experimental Results

In this section we present the results of several experimental runs conducted in the fields of central Pennsylvania. In Fig. 7, the plots show the ROC curve for human detection. There were nine runs with three different setups (which are listed in Table II).

TABLE II
EXPERIMENTAL SETUPS

| | Vehicle speed | # of runs | Total # of humans | |
|---|---|---|---|---|
| | | | moving | stationary |
| Uneven, off-road, bushes on side | 1 m/s | 1 | 1 | 0 |
| Pavement, bushes on side | 0 m/s | 4 | 4 | 0 |
| Pavement, tall grass on side | 5 m/s | 4 | 5 | 3 |

The pedestrians were walking or jogging. Clothed mannequins were used as stationary pedestrians. The sensor had a 17 Hz cycle rate. At each cycle it gives a new set of positive (pedestrian) and negative (poles, vegetation, ground, etc.) examples. On average there were about 350 positive examples and 5,000 (17,000 with ground not removed) negative examples per run. The ROC curve was produced by varying the threshold on the SOD.

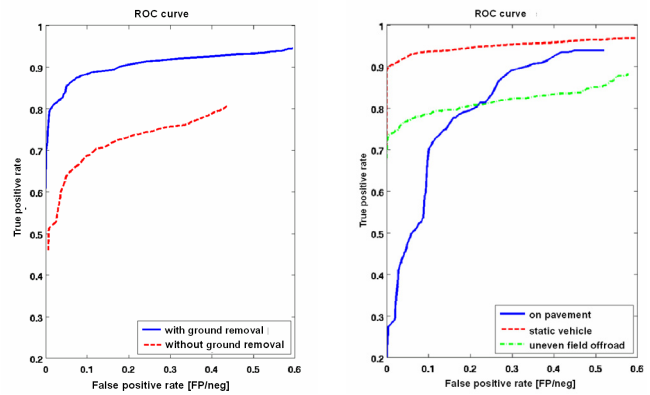


Fig. 7. Human detection performance. The ROC curves show the true positive rate vs. the rate of false positives per negative example. The plot was obtained using the threshold on the SOD as the dial variable. The left plot shows the improvement with ground removal. On the right the ROC curves for different setups are shown.

The left plot shows the combined ROC with and without ground removal. The improvement is quite significant, the true positive rate increases by about 0.15 on average. The ROC curve with ground removed shows the overall performance of the system. Table III shows the detection rates for three different false positive rates:

TABLE III
DETECTION RATES WITH GROUND REMOVAL

| False positive rate [%] | True positive rate [%] |
|---|---|
| 0.02 | 64 |
| 1 | 80 |
| 10 | 86 |

In Fig. 7, right side, the ROCs of three different setups are shown separately. The best performance is seen when the vehicle is stationary. In the off-road case the pedestrian often comes close to vegetation. The ROC curve for the runs on pavement is lower because of the stationary pedestrian. The system can not distinguish between a stationary pedestrian and a human-size fixed object (e.g. a pole or barrel). Other reasons why pedestrian were not detected with high confidence for some cycles is that a few cycles are needed to establish that an object is a human and sometimes the system loses the object and a track needs to be reacquired.

The detection range was limited by the range of the sensor. We detected humans at distances of up to 30 m.

## V. CONCLUSION

We have described a pedestrian detection and tracking strategy where a 3D LADAR is used to allow operation in uneven terrains, by first estimating the ground elevation, and then performing the detection using the measurements corresponding to a certain height above the ground. In contrast with other approaches, the information pipeline used to feed sensor data into our system is the same for both line and 3D scanners, which makes it possible to fuse data from multiple scanners. The system has been tested in the field, yielding improved detection results through the use of ground removal.

The main drawback of the current strategy is the inability to distinguish between a stationary pedestrian and a human-size fixed object. This issue is the focus of current research, where the exploitation of the information contained in the point cloud provides clues to discriminate against static objects.

## ACKNOWLEDGMENT

We would like to thank General Dynamics Robotic Systems for their support.

## REFERENCES

[1] K. Fuerstenberg and J. Scholz. Reliable Pedestrian Protection Using Laserscanners. *Proc. 2005 IEEE Intelligent Vehicles Symposium*, 6-8 June 2005, pp. 142 – 146.

[2] Y. Bar-Shalom, and T.E. Fortmann. *Tracking and Data Association*, Mathematics in Science and Engineering, Academic Press, 1988.

[3] L.E. Navarro-Serment, C. Mertz, and M. Hebert. Predictive Mover Detection and Tracking in Cluttered Environments. *Proc. of the 25th. Army Science Conference, November 27-30, 2006.*

[4] C.M. Shoemaker and J. A. Bornstein. The Demo III UGV Program: a Testbed for Autonomous Navigation Research. *Proc. of the IEEE Int. Symposium on Intelligent Control*, Gaithersburg, MD, September 1998, pp. 644-651.

[5] M. Hebert. Active and Passive Range Sensing for Robotics. *Proc. 2000 IEEE Int. Conf. Robotics and Automation*, vol. 1, pp. 102-110.

[6] Z. Sun, G. Bebis, and R. Miller. On-road Vehicle Detection: a Review. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Volume 28, Issue 5, May 2006. pp. 694 – 711.

[7] K. Fuerstenberg, K. Dietmayer, and V. Willhoeft. Pedestrian Recognition in Urban Traffic Using a Vehicle Based Multilayer Laserscanner, *Intelligent Vehicle Symposium*, 2002, pp. 31-35.

[8] D. Streller, K. Furstenberg, and K. Dietmayer. Vehicle and Object Models for Robust Tracking in Traffic Scenes Using Laser Range Images, *Intelligent Transportation Systems*, 2002, pp. 118-123.

[9] M. Lindstrom, and J. Eklundh. Detecting and Tracking Moving Objects From a Mobile Platform Using a Laser Range Scanner. *Proc. 2001 IEEE International Conference on Intelligent Robots and Systems*, 2001, pp. 1364-1369.

[10] A. Fod, A. Howard, and M. Mataric. A Laser-Based People Tracker. *Proc. 2002 IEEE International Conference on Robotics and Automation*, pp. 3024-3029.

[11] J. Cui, H. Zha, H. Zhao, and R. Shibasaki. Laser-based detection and tracking of multiple people in crowds. *Computer Vision and Image Understanding*, (106):300-312, 2007 .

[12] D. Morris, B. Colonna, and P. Haley. Ladar-based Mover Detection from Moving Vehicles. In *Proceedings of the 25th Army Science Conference*, November 2006.

[13] S. Thrun. Robotic Mapping: A Survey. *Technical report CMU-CS-02-111, Carnegie Mellon University, February 2002.*

[14] C.-C. Wang, C. Thorpe, M. Hebert, S. Thrun, and H. Durrant-Whyte. Simultaneous Localization, Mapping and Moving Object Tracking. *The International Journal of Robotics Research*, 26(9):889-916, 2007.

[15] A. Kelly and A. Stentz. Rough terrain Autonomous Mobility - Part 2: An Active Vision, Predictive Control. *Autonomous Robots*, No. 5, May, 1998, pp. 163 - 198.

[16] R. MacLachlan and C. Mertz. Tracking of Moving Objects from a Moving Vehicle Using a Scanning Laser Rangefinder. *Proc. 9th International IEEE Conference on Intelligent Transportation Systems*, Toronto, Canada, September 17-20, 2006.