In 26th Army Science Conference, Orlando, Dec. 2008.

LADAR-BASED VEHICLE DETECTION AND TRACKING IN CLUTTERED ENVIRONMENTS

Daniel Morris, Regis Hoffman, Steve McLean

General Dynamics Robotic Systems,
1501 Ardmore Blvd, Pittsburgh, PA 15221

Abstract

Detecting and tracking vehicles is crucial for safe operation of Unmanned Ground Vehicles (UGVs), but is challenging in cluttered, real-world environments. Here we present a method for discriminating vehicles from clutter found in natural terrain such as foliage, steep slopes, rock-outcrops, etc. Our method relies on a scanning LADAR and combines an obstacle detector and tracker, a vehicle modeling scheme, and a Support Vector-based discriminator. The output of our real-time system is a list of labeled obstacles and vehicles along with their positions, sizes and velocity estimates. This is used by a planner to enable autonomous navigation in the presence of other vehicles and significant clutter. We provide a quantitative analysis of the performance of our algorithm.

1. Introduction

There is an urgent need for autonomous and semi-autonomous vehicles to operate safely in real-world environments. A key step to achieving this is to determine the locations and trajectories of other vehicles in the vicinity of the Unmanned Ground Vehicle (UGV), and simultaneously to avoid falsely labeling other objects as vehicles. This task of detecting and discriminating stationary and moving vehicles from clutter in the vicinity of an unmanned platform is the goal of this work.

Nearby vehicle detection is required by the autonomous mobility system of a UGV. A sensor, typically a LADAR, provides an obstacle map of the world around the UGV. Then a predictor estimates likely states of the world in the near future, and the planner finds a trajectory towards the goal that avoids predicted obstacles. This task can be greatly simplified with certain world assumptions. The first is a static world model in which objects are important only to the extent that they obstruct the UGV motion (Lacaze *et al.* 2002). In this case identifying objects as vehicles is not important; just identifying which objects are impassible. The second simplified model is to include moving vehicles, but to eliminate most of the clutter using a road network map, such as the map provided for the 2007 Urban Challenge. Since non-vehicle obstacles are few, it is possible to use a conservative assumption that all obstacles are vehicles and still navigate well. The more difficult scenario, which we address here, is where there are vehicles (both moving and stationary) as well as significant clutter. In these cases, discriminating clutter from vehicles is crucial. If all clutter objects are treated as vehicles that may move, then the planner will be overwhelmed and unable to find a path that safely avoids collisions. To address this problem, our focus is on discriminating which objects in the vicinity of a UGV are clutter and which are vehicles.

Much of the successful autonomous navigation through static clutter relies on scanning LADARs (Langer *et al.* 1994; Lacaze *et al.* 2002; Matthies *et al.* 2003). LADAR data provide accurate range estimates and so

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **DEC 2008**	2. REPORT TYPE **N/A**	3. DATES COVERED **-**
4. TITLE AND SUBTITLE **Ladar-Based Vehicle Detection And Tracking In Cluttered Environments**		5a. CONTRACT NUMBER
		5b. GRANT NUMBER
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **General Dynamics Robotic Systems, 1501 Ardmore Blvd, Pittsburgh, PA 15221**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR’S ACRONYM(S)
		11. SPONSOR/MONITOR’S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES
See also ADM002187. Proceedings of the Army Science Conference (26th) Held in Orlando, Florida on 1-4 December 2008, The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**UU**	**7**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

can directly populate terrain and stationary obstacle maps which can be categorized based on appearance (Madhavan *et al.* 2004; Lalonde *et al.* 2006). Handling movers has proven difficult, although there are recent results for mover detection and tracking (Kluge *et al.* 2001; Wang *et al.* 2003; Morris *et al.* 2006; Morris *et al.* 2008) and mover prediction (Mertz *et al.* 2005; Navarro-Serment *et al.* 2006). Since vehicles may be stationary or move, we depend on 3D shape and not motion to discriminate vehicles from clutter. Rather than use edge features from a line-scanning LADARs as used in Keat *et al.* (2005) to find vehicles in parking lots, we use the full sampled surface for discrimination. We do not address the problem of human detection, as that is the focus of other work (Thornton *et al.* 2008). We use the tracker introduced in Morris *et al.* (2008) to help in the detection task, and Support Vector Machines (SMVs) (Joachims 1999) to learn a vehicle discriminator. Fig. 1 illustrates the type of cluttered scene in which we need to detect vehicles.

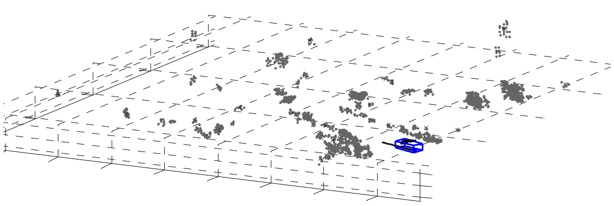


Figure 1. A sensor platform UGV (blue rectangle) moving along a road through wooded terrain. LADAR hits (after ground removal) are shown as grey dots. Using these LADAR returns, the UGV must determine the location of any vehicles in its vicinity.

2. Sensor and Platform

Our GDRS Generation IV LADAR has multiple lasers and time-of-flight detectors scanning a fixed pattern at roughly 10 Hz. The traversal of a cycle through this pattern we call a frame. If desired, multiple of these LADARS can be placed on a sensor platform to obtain 180 or 360 degree field of view coverage (see Fig 2). The LADAR data are coupled tightly with an INS-based navigation system enabling conversion of range data into 3D points.

Figure 2. Our UGVs with LADAR sensors

3. Algorithm Description

Given a high flow rate of 3D point samples of the world around the UGV, the challenge is to find all the vehicles and estimate their motion if they are moving. We structure this problem into three major components. The first step is filtering and clustering of 3D points. This is crucial to reducing complexity of the data association problem. By removing hits on the ground plane and by doing data association on clusters of points rather than raw points, the number of entities to search over is reduced by between 2 and 3 orders of magnitude. The second step is model fitting and tracking. Here data association is done leveraging the vehicle kinematics. Then the third step is vehicle discrimination. Each object track is analyzed to determine if it is a vehicle or clutter. These three steps are described in more detail in the remainder of this section.

3.1 Filtering and Clustering of 3D points

The first filtering step is to remove hits on the ground surface. We have several techniques that work similarly well: growing the ground surface radially outwards with thresholds on slope, or fitting roughly horizontal planes in a faceted manner over the scan, (Morris *et al.* 2006). This reduces the data flow rate and provides a spatial separation of points belonging to different objects.

The next step is a segmentation of the points into objects or clusters. We have used a variety of methods with good success including mean-shift as in Morris *et al.* (2006) and simple 2D binning with by local maxima estimation and region growing. The important requirement is that it provides an over-segmentation of the data and that each cluster belongs to at most one vehicle (or object). In the later model-fitting stage, clusters from the same object will be merged.

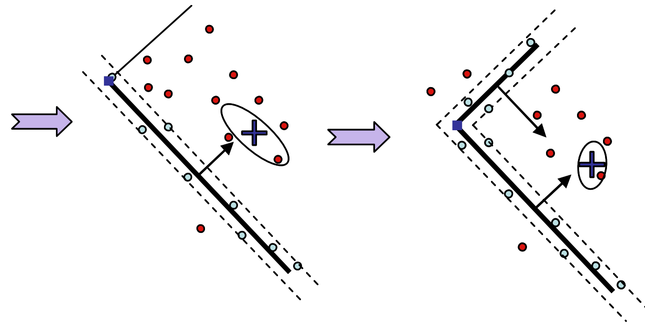
3.2 Model Fitting and Tracking

The next step is to create vehicle hypotheses; that is, possible vehicle locations and poses, each of which will be evaluated in Section 3.3. Each hypothesis starts out as one or more clusters of points and is refined to an oriented rectangular region representing a vehicle with its position, pose and size.

Now the assignment of clusters to hypotheses is potentially computationally expensive. For example, considering all pairs and all triples of clusters as possible hypotheses quickly becomes unmanageable. To avoid this complexity, we use the following greedy assignment algorithm:

1. For all existing tracks, create hypotheses at predicted locations and assign clusters at those locations to those hypotheses. If there is a conflict over a cluster, assign it to track with highest probability of being a vehicle at that location.
2. For unclaimed clusters, starting with closest and proceeding in range order, assign it to a new hypothesis and then:
    a. Fit our vehicle shape model to all the points (see below).
    b. If the shape model overlaps unclaimed cluster centroids, add them to the hypothesis and repeat step 2.a.

As part of hypothesis creation, we estimate a vehicle's position and orientation. To do this we assume a vehicle has a roughly rectangular exterior shape (viewed top-down) of which our sensor will observe one or two sides. We then robustly fit an L-shape or a single edge (if only one face is visible) to all the points in the hypothesis as illustrated in Fig. 3, and more details can be found in Morris *et al.* (2008). This is similar to 2D fitting done in Keat *et al.* (2005); Wang *et al.* (2003). Also, we are able to estimate vehicle dimensions, although this is done over a series of frames to avoid including clutter.



(*a*) (*b*)

Figure 3. Robust data fitting of an edge (thick dark line) (*a*), or an "L-shape" (*b*). The inlier LADAR points (shown in green) are used to improve the fit, which in each case defines a corner position and orientation. Using this corner, the vehicle center is estimated along with a covariance.

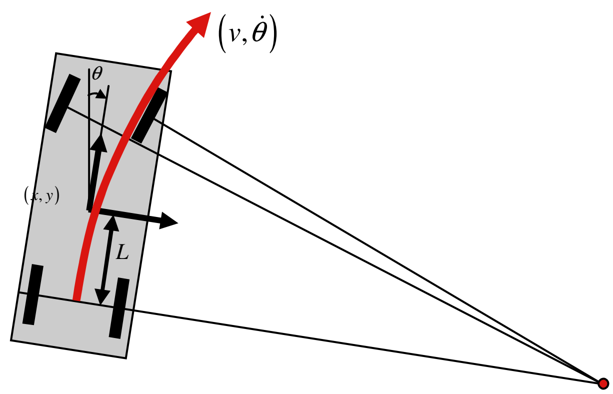


Figure 4. Our VASM kinematic model for vehicle tracking constrains motion to be perpendicular to the axis of rotation, whose distance from the vehicle center, *L*, is estimated by the filter.

We treat the robust vehicle-model fit as a measurement of position, (x,y), and pose, θ. Using this measurement model, a state vector, $\boldsymbol{x}$, and transition matrix $\boldsymbol{\Phi}(\boldsymbol{x})$, we have the essentials for a Kalman Filter-based tracker. The state transition is governed by the vehicle kinematic model. To model Ackerman steering, as well as other steering models, we developed a kinematic model which we call the Variable Axis Steering Model (VASM), first introduced in Morris *et al.* (2008). The state vector is: $\boldsymbol{x} = \begin{pmatrix} x & y & L & v & \theta & \dot{\theta} \end{pmatrix}^T$.

The vehicle or object proceeds with speed, v, along an arc tangent to the vehicle orientation. The distance from the vehicle center, L, of the axis of rotation, is estimated as one of the parameters, see Fig. 4. Further details of the state transition matrix are in the appendix. Our tracker consists of a multi-hypothesis Kalman Filter that takes measurements from the robust vehicle fitter and predicts motion with the VASM kinematic model.

3.3 Vehicle Discrimination

To this point we have a detector and tracker for any large object. From an autonomy perspective, vehicles need to be treated differently than other objects as they have potential to move into our planned trajectory. Hence the next step is to discriminate which of the tracked objects are vehicles.

There are a number of factors that make discriminating vehicles from clutter challenging. The primary one is the variable resolution and sampling of the 3D points on the object surface. Also, as range increases the number of hits falls off with the square of the range, making the need for low-resolution discrimination important. In addition, the 3D appearance of a vehicle varies depending on viewing perspective, self-occlusions, the surface reflectivity, grazing angle, the LADAR noise and its interaction with surface reflectivity. For example, some shiny surfaces give no returns at shallow grazing angles, and the difference in returns from a shiny and a matte black surface can lead to differences in depth estimation. Given all of these factors it is difficult to create an *a priori* generative model for vehicle appearance. Instead, our approach is to develop a discriminative model that can be trained on actual LADAR data of both vehicles and clutter.

The tracker described in section 3.2 provides two very useful functions for the discriminator. It groups clusters into a single object. Also, it provides a position and orientation estimate, and hence the alignment of the 3D points onto a local coordinate system fixed on the vehicle. It thus acts as an interest operator providing the pose and location of a vehicle hypothesis.

Our feature space consists of a projection of the 3D points into a 3D grid positioned in the local coordinate system. The coordinate system is aligned with the corner of the object or vehicle being tracked. The side of the vehicle hypothesis is oriented along the positive X axis and the front or rear along the positive Y axis. (When the front right or rear left corners are tracked, the points are reflected across the X axis to fit this model.) We call this projection a binned density model, which we represent as a normalized vector in a high dimensional space. An example of this model creation is illustrated in Fig. 5.

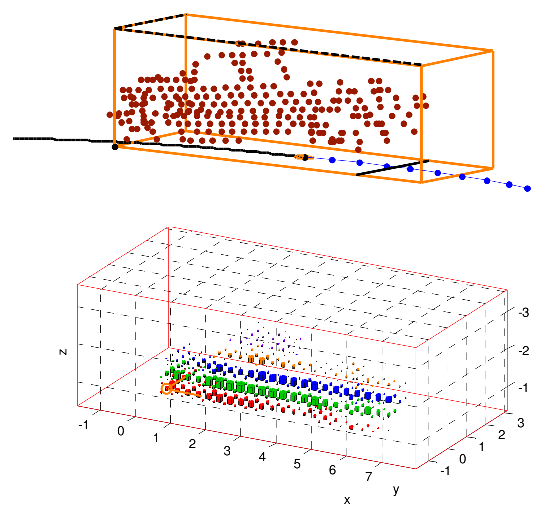


Figure 5. Vehicle model creation: the LADAR hits from a tracked vehicle (top) are converted into a binned model (below). In this case the front-left corner is being tracked and this used to align the data points before binning. This binned density is normalized and used as a high-dimensional feature vector for classification.

Now the end product of filtering, clustering, model fitting, tracking and binning is a set of feature vectors representing each vehicle hypothesis. We use SVMs (Joachims 1999) to learn a discriminative model for separating vehicles from clutter. For simplicity, and to avoid over-fitting, we have limited ourselves to linear SVMs. However, vehicle shape appearance depends on range both because of sampling resolution and because the LADAR is on the roof of the sensor platform, and so looks down on close-by vehicles, but horizontal to long range vehicles. Hence we investigated using separate classifiers for different ranges.

4. Experiments

Training and testing requires labeled data. With moderate care in data collection, it is possible to almost fully automate the labeling. Clutter objects are collected and labeled by simply driving the sensor platform through scenery with no other vehicles, and tracking all the target clusters. Vehicle data are collected by driving a target vehicle in front of the stationary sensor platform. The tracker will detect and track the mover which is known to be a vehicle so can be automatically labeled. Data containing clutter and stationary vehicles requires some manual labeling, but is greatly aided by the tracker.

We collected a large volume of clutter data by driving our sensor platform along trails through sparse and dense vegetation and over a variety or terrain. For vehicles we collected data on large pickups, mid-sized sports utility vehicles and our small XUV robotic vehicle. Our test runs included usual traffic scenarios such as at intersections and along roads, and driving along narrow roads with significant clutter. Our algorithm was tested both on stored data and real-time data with output going to the autonomy planner.

Training was performed on 10,000 positive examples and 20,000 clutter examples, and similar quantity of separate data was used for testing. Rather than training a single model, we obtained improved performance by training 3 models for 3 different target ranges: under 20m, 20 to 40m, and 40 to 60m. Our vehicle model has a horizontal resolution of 30cm and vertical of 40cm forming a grid of length 32, width 16 and height 8 giving a feature vector with 4096 dimensions. The discriminator trained in this space is illustrated in Fig. 6. We experimented with adding two additional components to this vector: a score between 0 and 1 indicating how evenly the object edge points are distributed on the visible edges, and an orientation measure indicating if the target vehicle is parallel or perpendicular to the viewing ray.

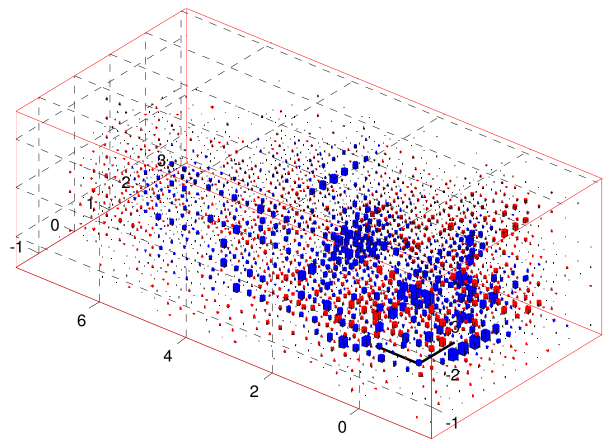


Figure 6. The result of training is a separating hyperplane in feature space. This is illustrated here with blue rectangles representing positive values and red negative. As can be seen, vehicle features close to the corner (marked in black) are emphasized.

5. Results

Our algorithm runs in real time on a standard Pentium Core 2 Duo handling up to about 100 targets at 10 Hz. Fig. 7 illustrates the benefits of integrating the discriminator in the tracker. Vehicle hypotheses that pass a minimum fit-score are shown as rectangles, and our discriminator eliminates a great majority of these as non-vehicles. A quantitative performance analysis of our discrimination algorithm at different target ranges is shown in Fig. 8. We see good performance up to about 40m, beyond which the declining resolution leads to higher misses and mistakes. One surprising observation is that the best performance is between 20 and 30m, and closer-in the performance is poorer. The reasons for this are unknown; possibilities include feature alignment being poorer, or that bushes and other clutter, when observed by a close-in LADAR, more closely approximate a vehicle shape.

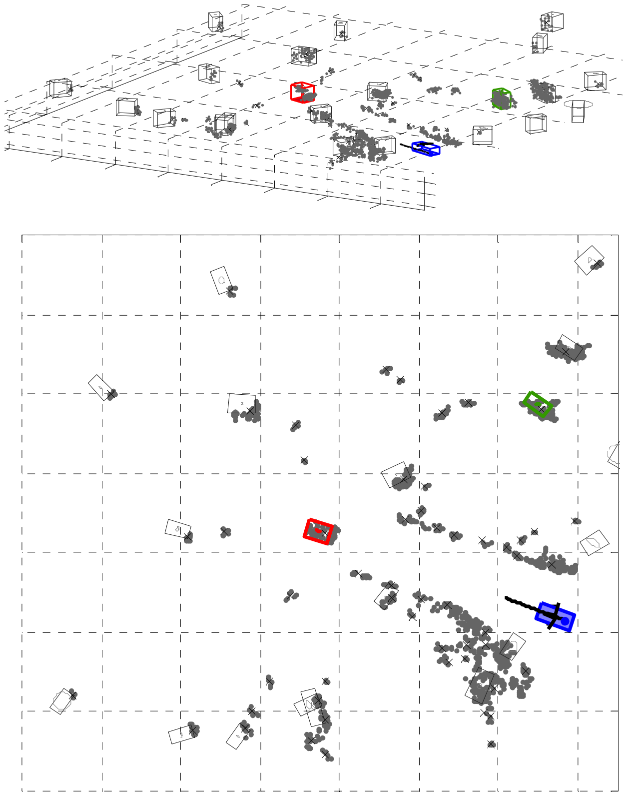


Figure 6. The same wooded scene as in Fig 1, shown in 3D and top-down. Vehicle hypotheses that pass a minimum fitting threshold are shown as rectangles. Our discriminator identifies all but two of these as clutter. The red rectangle is a true vehicle and the green is a bush or tree that appears similar to a vehicle in shape.

6. Conclusion

We developed an easily trainable vehicle detector that uses scanned 3D shape alone to discriminate vehicles from clutter. By tightly integrating this discriminator into the tracker we are able to detect and track vehicles in high-clutter, natural environments from a moving UGV.

There are limits to discrimination from LADAR data alone, particularly at longer ranges and in urban environments. Objects like Jersey barriers can appear from some views very similar to a vehicle, even to a human, and especially at long range due to the low resolution of the LADAR. To address this, we plan to increase resolution by integrating data temporally. We are also working on fusing results with other sensing modalities.

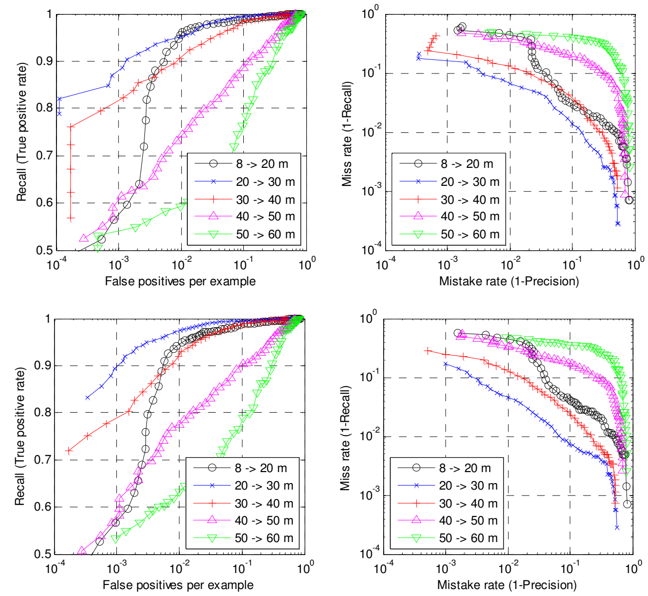


Figure 7 Performance of our discriminator. On the left are ROC charts showing ability to filter clutter at various ranges. On the right are the Detection Error Tradeoff charts showing the miss rate versus mistake rate. The top row shows results for just binned density features. Below this are results when fitting-score and estimated-relative-orientation are included in the feature vectors.

Acknowledgements

This work was sponsored by the Army Research Lab under contract: DAAD 19-01-2-0012.

References

Joachims, T., 1999. “Making large-Scale SVM Learning Practical. Advances in Kernel Methods - Support Vector Learning”, B. Schölkopf and C. Burges and A. Smola (ed.), MIT-Press.

Keat, C., Pradalier, C., Laugier, C., 2005. “Vehicle detection and car park mapping using laser scanner,” Proc of Intelligent Robots and Systems (IROS-05), pp. 2054-2060.

Kluge, B., Kohler C., Prassler, E. 2001. “Fast and Robust Tracking of Multiple Objects with a Laser Range Finder, IEEE Int. Conf. on Robotics and Automation,” pp. 1683-88.

Lalonde, J., Vandapel, N., Huber, D. Hebert, M., 2006. “Natural Terrain Classification using Three-Dimensional Ladar Data for Ground Robot

Mobility," in Journal of Field Robotics, Volume 23, Issue 10.

Lacaze, A., Murphy, K., DelGiorno, M. 2002. "Autonomous mobility for the Demo II experimental unmanned vehicles," in Proc. Assoc. Unmanned Vehicle Systems Int. AUVSI-02.

Langer, D., Rosenblatt, J., and Hebert, M., 1994. "A Behavior-Based System for Off-Road Navigation," IEEE Trans. Robotics and Automation, Vol. 10, No. 6, December, 1994, pp. 776-782.

Madhavan, R., Hong, T., 2004. "Robust Detection and Recognition of Buildings in Urban Environments from LADAR Data," Proc. Applied Imagery and Pattern Recognition, AIPR-04.

Matthies, L., Bergh, C., Castano, A., Macedo, J., Manduchi, R., 2003. "Obstacle Detection In Foliage with LADAR and Radar" in International Symposium on Robotic Research.

Mertz, C., Duggins, D., Gowdy, J., Kozar, J., MacLachlan, R., Steinfeld, A., Suppe, A., Thorpe, C., Wang. C., 2005. "Collision Warning and Sensor Data Processing in Urban Areas," Proc. 5th Int. conf. on ITS telecoms., pp. 73-78.

Morris, D., Colonna, B., Haley, P., 2006. "LADAR-based Mover Detection from Moving Vehicles," in 25th Army Science Conference.

Morris, D., Haley, P., Zachar, W., McLean, S., 2008. "LADAR-Based Vehicle Tracking and Trajectory Estimation for Urban Driving", in Proc. Assoc. Unmanned Vehicle Systems Int. (AUVSI-08), San Diego.

Navarro-Serment, L., Mertz, C., Hebert, M., 2006. "Predictive Mover Detection and Tracking in Cluttered Environments," in Proc. 25th Army Science Conference.

Thornton, S., Hoffelder, M., Morris, D., 2008. "Multi-sensor Detection and Tracking of Humans for Safe Operations with Unmanned Ground Vehicles," Workshop on Human Detection from Mobile Platforms, Pasadena.

Wang, C., Thorpe, C., Suppe, A., 2003. "LADAR-based Detection and Tracking of Moving Objects from Ground Vehicle at High Speeds", Intelligent Vehicles Symposium, pp: 416 – 421.

Appendix

Here we define the state transition matrix, $\boldsymbol{\Phi}(\boldsymbol{x})$, of the VASM kinematic model from Morris *et al.* (2008). Define a local coordinate system located at the vehicle center at time t_0, and with its x-axis aligned with the vehicle orientation, θ. The vehicle center $\left({}^c x, {}^c y\right)$ moves in an arc defined in these local coordinates:

$$
\begin{aligned}
{}^c x\left(t_0+\Delta t\right) &= v\Delta t \operatorname{sinc}\left(\dot\theta\Delta t\right) + 2L\sin^2\left(\dot\theta\Delta t/2\right) \\
{}^c y\left(t_0+\Delta t\right) &= v\dot\theta\Delta t^2 \operatorname{sinc}^2\left(\dot\theta\Delta t/2\right)/2 - L\sin\left(\dot\theta\Delta t\right)
\end{aligned}
$$

Using this, the state transition matrix is:

$$
\boldsymbol{\Phi} = \boldsymbol{I} + \boldsymbol{T}\begin{pmatrix}
0 & 0 & \partial\,{}^c x/\partial L & \partial\,{}^c x/\partial v & -{}^c y & \partial\,{}^c x/\partial\dot\theta \\
0 & 0 & \partial\,{}^c y/\partial L & \partial\,{}^c y/\partial v & {}^c x & \partial\,{}^c y/\partial\dot\theta \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \Delta t \\
0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}
$$

where $\boldsymbol{T}$ contains a 2D rotation, $R(\theta)$, back into world coordinates:

$$
\boldsymbol{T} = \begin{pmatrix} R(\theta) & 0 \\ 0 & \boldsymbol{I}_4 \end{pmatrix},
$$

and the partials with respect to each parameter are:

$$
\frac{\partial\,{}^c x}{\partial L} = 2\sin^2\left(\dot\theta\Delta t/2\right), \qquad \frac{\partial\,{}^c y}{\partial L} = -\sin\left(\dot\theta\Delta t\right)
$$

$$
\frac{\partial\,{}^c x}{\partial v} = \Delta t \operatorname{sinc}\left(\dot\theta\Delta t\right), \qquad \frac{\partial\,{}^c y}{\partial v} = \dot\theta\Delta t^2 \operatorname{sinc}^2\left(\dot\theta\Delta t/2\right)/2
$$

$$
\frac{\partial\,{}^c x}{\partial\dot\theta} = v\Delta t^2 \operatorname{sinc}'\left(\dot\theta\Delta t\right) + L\Delta t\sin\left(\dot\theta\Delta t\right)
$$

$$
\begin{aligned}
\frac{\partial\,{}^c y}{\partial\dot\theta} = {} & v\Delta t^2\left(\operatorname{sinc}\left(\dot\theta\Delta t/2\right)\cos\left(\dot\theta\Delta t/2\right) - \operatorname{sinc}^2\left(\dot\theta\Delta t/2\right)/2\right) \\
& -L\Delta t\cos\left(\dot\theta\Delta t\right)
\end{aligned}
$$