An Obstacle Representation for Off-road Autonomous Driving

Joseph Weber          Larry Matthies

Electrical Engineering Department
The California Institute of Technology

The Jet Propulsion Lab
jweber@vision.caltech.edu    lhm@robotics.nasa.jpl.gov

Abstract

The field of autonomous vehicles has seen considerable success in the domain of freeway driving. Here the constraints of freeway construction, rules of the road, and encoded maps are used to successfully drive a vehicle with a minimum of human intervention. Autonomous navigation in a more general environment has seen less progress however. The off-road autonomous system must be able to identify both potential obstacles of any form as well as determine which areas in view are safely traversable.

The most common sensing modality used for mobile vehicles is range information. Range information is used to identify potential hazards, traversable regions, and possible landmarks. However, obstacle detection algorithms invariably produce some false alarms, even with relatively good range data. The accuracy of obstacle detection can be improved by exploiting temporal information. This requires efficient methods to represent and track detected obstacles over time.

We propose an object representation which allows the integration of range information with time, without the large computational and storage requirements of earlier representations. Potential obstacle regions in the image are described in terms of their range and direction with respect to the vehicle. Much effort has been put into matching Cartesian maps from frame to frame given the large uncertainty for distant objects. A directional representation however maintains the high accuracy of directional information (which is limited only by pixel density and camera calibration). Tracking objects from frame to frame is easier since the potentially large uncertainty in range does not corrupt the directional information.

The use of this representation to identify untraversable regions is demonstrated on a series of real images taken from an off road vehicle.

1 Introduction

Autonomous navigation of a robotic vehicle has been one of the first tasks investigated by machine vision researchers. The past decades have seen much research, with various degrees of success. A significant amount of success has been seen lately in the realm of highway driving, where the constraints of highway construction and rules of the road are exploited to reduce computation [DM92].

For less structured environments, more computational resources must be used to identify the larger degree of variability in obstacles and terrains. In off-road driving, obstacles may be of any size and shape. A representation must be used which can handle this variability.

Many of the approaches to unstructured environments use range imagery obtained from either passive or active vision systems. From this range data, an attempt is made to find traversable and untraversable regions of the visible field. This is often done through the use of a grid-based representation. Examples are occupancy grids or elevation maps [El89, EM87, Nii84, LRH94, KK92, BK91]. Both use an array of locations in front of the vehicle which contain information about either the terrain or probability of an obstacle existing at that location.

The use of grid-based representations has a number of limitations. We enumerate some of them here.

1. Two or three dimensional grids often require the storage of a large amount of data. Depending on
the resolution of the grid, the number of elements in the grid can be large.

2. Related to the first problem is the problem of registration between frames. Because of the large amount of data in the grid representation, finding the correspondence between grids when the vehicle moves can be computationally expensive. In addition, due to the variation in reliability of measurements within each grid location, correspondence is subject to error.

3. Cartesian grids can be inefficient representations of range data. Cartesian maps often have large regions of missing data formed from range shadows: locations in the scene not visible from the camera because of occluding objects. As a result, large parts of the grid might be labeled as "unknown".

4. Finally, the use of grids often does not take into account the nature of the uncertainty in range measurements. The uncertainty in range grows quadratically with distance. The angular uncertainty is fixed by the size of pixels used. When range measurements are placed in Cartesian grids, the uncertainty is spread non-uniformly into all three dimensions.

We propose a representation that addresses these limitations, at the cost of some simplifying assumptions. In our model, regions of the range image which are identified as potential objects are projected onto a plane parallel to the ground surface. In this representation each object is represented by two lists of numbers: the minimum and maximum range at each angular direction within the object's angular span. This is illustrated in Figure 1. Each column in the image corresponds to a single radial direction measured from the camera's vertical. We collapse the range measurements within this column into the maximum and minimum values. As a result, each object can be represented by a small number of measurements. The maximum amount of data storage is two values times the angular resolution of the representation.

This representation addresses each of the limitations enumerated above: As stated in the above paragraph, each object is represented by a limited set of values. By using an object based representation we simplify the task of finding correspondences between frames. Based on the motion of the vehicle, object locations are predicted and then matched with new measurements. This is similar to the predict and match methods of Kriegman et al [KTB89] which used line segments instead of regions. Range shadows are explicitly modeled as continuous areas between measurements. The angular representation leaves the uncertainty in the measurements along the radial direction and does not mix the uncertainty with the angular resolution which is quite precise (on the order of pixel size). The planar representation is similar to Miura and Shirai [MS94] which also used only line segments as the object representation.

In the next section we outline how obstacles are detected from the range image. In Section 3 the representation in terms of planar blobs is explained in more detail, including the dynamic update of estimates of the blobs. Finally, in Section 4 we show results on images taken from an off-road vehicle.

![Figure 1: The two obstacles in the range image on the left are mapped into the planar view on the right. A column in the range image is condensed into two values in the planar view: the minimum and maximum range.](image)

2 Obstacle Detection

Obstacles in the path of the vehicle are determined solely from the range data. We use two methods for labeling parts of the image which could be obstacles. Both methods use criteria based on pixel level data only. The criteria are based on detectable differences in the image itself, independent of how the images were obtained. For example, a detection threshold might be 0.5 pixels instead of 10cm. Using camera calibration parameters, this pixel-level threshold can then be translated into real-world metrics.
2.1 Obstacles from a planar assumption

The first method for potential obstacle detection is based on the assumption that there exists a traversable region immediately in front of the vehicle and that this region is, to first order, planar in world coordinates. We use the range data from the immediate vicinity of the vehicle to find a planar approximation. Variations from this plane are potential obstacles.

The stereo disparity for cameras with parallel optical axes looking at a planar surface is affine in image coordinates, i.e. the disparity \( d \) is a function of image coordinates \( x \) and \( y \) via \( d = Ax + By + C \). From the disparity image of a scene containing a plane, an affine fit gives an estimate of the coefficients \( A, B, C \). Given camera calibration parameters, there exists a one to one correspondence between the coefficients \( A, B, C \), and the height and orientation of the plane with respect to the camera.

The planar approximation is found using a robust least squares fit of the disparity. Disparity estimates which are more than one pixel away from the affine parameterization (outliers of the affine model) are labeled as potential obstacles. This corresponds to points which are above or below this reference plane. The threshold is in pixel units as opposed to a distance metric. A threshold in units of pixels is used since the noise in disparity is in units of pixels and assumed uniform throughout the image, whereas the uncertainty in range estimates is known to vary throughout the image. The result of a uniform disparity threshold is that the height of detected objects grows with range. This reflects the growing uncertainty in range measurements with range.

We track the orientation of the planar region in front of the vehicle from frame to frame using the known dynamics of the vehicle and the on-board inertial sensors. In this way, only a small correction to the robust fit is needed at each new frame. The algorithm is similar to the planar estimation method used in previous highway driving work [WKLM95].

2.2 Obstacles from range discontinuities

The second method for obstacle detection looks for depth discontinuities within the range data. These discontinuities would indicate rapid changes in depth due to occluding surfaces. This method specifically looks for so called “negative” obstacles, regions of the terrain that are within a range shadow, and thus not directly visible to the cameras. Such regions are depicted in Figure 2.

Given three range estimates (not necessarily adjacent in the image), we can form a line in three dimensional space between two of them. If the third point does not lie on that hypothetical line, it could indicate the presence of a range discontinuity. (Note that the figure shows a horizontal line between points. The two points which form the hypothetical line can however be at any orientation with respect to the camera). The distance, \( D \), from the hypothetical line as a function of range estimates \( y_i \) and pixel numbers \( x_i \) is

\[
D = \frac{\frac{y_1y_3(x_1 - x_2)}{y_1x_1 - y_3x_3} - y_2}{y_1x_1 - y_3x_3 - (y_1 - y_3)x_2}
\]

The value of \( D \) is actually computed in disparity space. In this way the disparity values do not need to be converted to range values. Also, the noise in disparity is uniform in the image and thus we can compare the value of \( D \) in disparity space to the expected noise in disparity. Obstacles are regions where the value of \( D \) in disparity space is above the expected noise.

For points near the vehicle, the expected noise variance in the range estimates can be very small. The magnitude of detectable distances \( D \) by the above criterion might not be large enough to be considered an obstacle. At this point, a real-world metric is used to determine if the discontinuity is large enough to be considered an obstacle.

3 Obstacle Representation

As described in the introduction, each detected obstacle region in the image is represented as a blob in a planar projection onto the driving plane. This is performed by scanning each column of the image containing the obstacle and finding the minimum and maximum range of pixels in that column. A single column in the image corresponds to a plane with a unique orientation with respect to the camera’s vertical. That is, all image pixels in column \( c \) of the image correspond to direction rays at an angle \( \theta = \arctan((c - c_0)/f) \) from the central vertical plane through the optical axis. \( c_0 \) is the column containing the optical center, and \( f \) is the focal length in pixels. This assumes of course a pinhole model for the camera.
Figure 2: Negative obstacles are detected if the ratio of the distance \( D \) divided by the expected uncertainty in range is greater than a threshold. The ratio is scale independent. It is also easier to detect than the value of \( H \) which is sensitive to camera calibration.

In projecting the obstacles to a plane we discard information on the variation in range with height. We also do not keep information on the actual height of the obstacle. Since all obstacles are regions the vehicle should not drive over, it is not important to know the height of the obstacle. Secondly, the representation is convex in range. The minimum and maximum range are single valued functions of direction angle.

To form obstacle representations, we first group together all pixels in the image that were detected as potential obstacles, as outlined in the previous section. Pixels are grouped together if 1) they are connected in the image and 2) the difference in their range is less than the expected uncertainty in range. Any connected group consisting of more than 20 pixels is considered a potential obstacle.

3.1 Model Prediction

The locations of objects found in the image are translated to the next frame given an estimate of the vehicle’s motion between frames. For our experiments this information came from a differential GPS receiver on the vehicle. The prediction is formed by transforming each object in the planar representation to the new camera location. In addition to translating each object according to the motion, the uncertainty in object position is increased due to the uncertainty in the vehicle’s motion estimate.

Parts of the object which are predicted to be behind the vehicle are removed from local consideration. If the object just passed had significant confidence (see Section 3.2), it is added to a global map.

Objects found in the new image are identified as being either 1) associated with previously detected objects or 2) potentially new objects. Objects are associated with previous tracks if they significantly overlap in the planar representation space. Calculating this planar overlap is computationally simple. If a new objects is associated with an existing track, the representation of the tracked object is updated using the new information. This update takes into consideration the uncertainty of the both the old and new measurements. The uncertainty decreases with each new measurement.

Our experiments used only the planar motion components (planar translation and rotation about the vertical) of the vehicle. This is valid for relatively flat terrain and when the vehicle’s camera has limited tilt and roll. If the camera motion includes significant tilt and roll, the model would have to include these terms in the motion of objects.

3.2 Model Uncertainty

The range estimates are often noisy. We assume whatever sensor is providing the range data also provides an estimate of the uncertainty in that data. Object representations are updated from frame to frame based on the uncertainty in each measurement. The minimum and maximum range values are updated based on the relative uncertainties of the new measurement and the current model. This is similar to recursive estimation via Kalman filtering. A similar example is [KTB89].

The longer an object is tracked in the image, the smaller the uncertainty in model parameters. Those objects which have small enough uncertainty are added to a global map when the vehicle passes them. This global map can be used to identify landmarks for localization.

4 Experiments

We ran the algorithm on real sequences taken from a military jeep equipped with stereo cameras. The range maps were calculated in real time using the stereo vision algorithm of the JPL rover [Mat92]. In Figure 3 we show an outdoor scene with a rock in the road and a wall of rocks forming the road boundary. To the right is the over 5000 pixels identified as potential obstacles projected on to the ground plane.
and viewed from above the vehicle. These points were identified as belonging to two separate obstacles, each represented by fewer than 180 values (min and max range at one degree intervals).

In Figure 4 we show a sequence of images as the vehicle approaches a ravine. The left figures show the image with the detected obstacles outlined. The estimated horizon due to the planar approximation of the region in front of the vehicle is also shown. To the right of each image is the local object representation. Intensity is proportional to confidence that the object is truly an obstacle. At first, only the foreground edge and the back of the ravine are visible. The front edge is detected by the discontinuity detector while the back of the ravine is detected as being below the planar surface. Notice in the middle image of the figure that there is a large region of the ravine not visible in the range image because of the range shadow. This would be a significant problem for some occupancy grid methods.

5 Conclusion

A radial, planar representation for tracking potential obstacles to an off-road vehicle is presented. The representation was developed to overcome some of the limitations with grid-based methods. Experiments with real-world images demonstrated that objects could be represented with a limited number of parameter values. The representation promises to make tracking and estimation of object locations easier and thus facilitate the path planning and obstacle avoidance components of off-road driving. We feel that by tracking obstacles from frame to frame, the accuracy of obstacle detection can be increased. Range sensors often produce noisy data that prescribes the need for using temporal information to increase accuracy.

References


Figure 3: Dirt road scene. The image contained over 5000 pixels detected as being obstacle (projected in the middle image). Two connected objects were found and are shown in the right image. (For scale, the square in the lower part of the image is 2m on a side.)

Figure 4: Hidden ravine sequence. The vehicle is approaching a ravine. The left figures show the image with the detected obstacles outlined. The estimated horizon due to the planar fit of the region in front of the vehicle is also shown. To the right of each image is the local representation. Intensity is proportional to confidence that the object is truly an obstacle. At first only the forward edge and back of the ravine is visible. False obstacles have either lower confidence or are not tracked at all, since an object must be visible for at least two frames to be tracked.