Linking Sensed Images to an Ontology of Obstacles to Aid in Autonomous Driving

Craig Schlenoff

National Institute of Standards and Technology (NIST)
100 Bureau Drive, Stop 8230, Gaithersburg, MD 20899
craig.schlenoff@nist.gov

Abstract

In this paper, we discuss the importance of recognizing and representing both stationary and moving obstacles for the purpose of autonomous driving as well as linking these representation to an ontology of obstacles to aid in deducing additional information about them. With the ability to access additional information about a sensed obstacle, an autonomous vehicle can better forecast where that obstacle can and can not be at a future time, and therefore be able to better plan its path to avoid collision with that obstacle.

This paper describes work just recently begun at the National Institute of Standards and Technology in developing and incorporating an ontology of moving obstacles into the control of an autonomous vehicle to aid in path planning and obstacle avoidance.

Introduction

For the purpose of collision forecasting and avoidance in an autonomous vehicle, it is not only important to know where an object is at a given time, but also to accurately predict where that object will be at a given time in the future [Shoemaker99]. If the object is stationary, this is easy. However, if the object is moving, this becomes a lot more difficult. One must sense the existence of the object at a given time, and be able to correlate where that same object is in a sensed image at a time in the future. Once this is done over a series of time points, one could attempt to associate a set of equations with that motion as a function of time, and be able to predict where that object will be at a time in the future.

However, just knowing the direction and speed of an object at any given time does not necessarily indicate that the object will continue moving at that direction and speed in the future. Although there is no accurate way to definitively predict where a moving object will be at any point in the future, there are ways of narrowing down the possibilities with a high degree of certainty. For example, if one could identify that a particular moving object was a car, then one may know that the maximum speed of a car was 192 km/hr and the average speed of a car on a highway was 104 km/hr in open driving conditions. They may also know that the car can only move forward and backwards (not sideways) and could turn at a maximum angle given its speed and tire position. They may also know that cars primarily drive on roads, and stay on the roads, and within the lanes of the road, if at all possible. With this information, the planner for the car can eliminate locations that the car cannot possibly reach at a given time, and can also assign probabilities, given the known characteristics of the car, of where it is most likely to be at a given time. Perhaps the highest probability of where it will be is directly along its current path, slightly lower probability in the areas surrounding its current path, and zero probability in areas that it does not have the capability of reaching given its driving limitations.

Almost everything mentioned above requires that there be a mechanism to identify that the object being sensed can be classified as a car. In other words, there needs to be a mechanism to associate a set of pixels (or other sensor's representation), or a map-based representation, to a symbolic entity that represents the car.

Representing Objects

Object recognition is a very difficult challenge, and one that is only lightly glossed over in this paper. However, in the case of on-road driving, the challenge becomes a little bit easier since, for the most part, the environment is controlled and there is a bounded set of objects that a driver would

---

No approval or endorsement of any commercial products in this paper by the National Institute of Standards and Technology is intended or implied. This paper was prepared by United States Government employees as part of their official duties and is, therefore, a work of the U.S. Government and not subject to copyright.

expect to sense (e.g., other vehicles, pedestrians, animals, traffic lights, signs, road, etc.).
Since a single object can be sensed by many different sensors at the same time (e.g., stereo cameras, LADARs, FLIRs, etc.), the more information that can be captured outside of the representation of any single sensor, the better the probability that the information can be reused. For example, a vehicle six meters ahead could be represented by a set of points in a LADAR's representation, could be represented as a set of pixels within a camera's representation, and could be represented as a set of heat images within a FLIR representation. Instead of having to deal with each sensor separately, it is customary to use a map-based representation to fuse the various sensor images together [Hong02]. However, this map-base image is still primarily grid-based, with little ability to incorporate detailed attributes and rules that pertain to groups of grid cells that comprise a single object. For example, if that vehicle were detected to be moving, one would not want to associate that motion with every grid cell that comprised that vehicle. Instead, one would want to capture that information once, perhaps in a symbolic representation of a vehicle, and link the portions of the map-based representation to that symbolic representation. In addition to representing the vehicle's motion, there may be some other pertinent information about the vehicle that would be beneficial to represent symbolically, such as information to detect where the vehicle may be at a time in the future, the cost of colliding with the vehicle, and information to help recognize the vehicle. Examples of these types of information are shown below:

For the purpose of recognizing the vehicle from sensed data:

- Various salient features of the vehicle, such as existence of important structures, shapes, lines and curves, and their relationship to one another
- Pointers to CAD drawings showing prototypical pictures of vehicle (see Figure 1)
- General knowledge, such as the fact that cars are typically found on roads

Facts for the purpose of location prediction:

- A car typically has a maximum forward velocity of approximately 192 km/hr and a maximum backwards velocity of 64 km/hr. However, a car typically moves forward at a speed of between 8 and 120 km/hr.
- A car can only move forward and backward, but can also turn at a maximum angle of 45°.
- The angle that a car can turn is usually a factor of its speed.
- Associated equations capturing the current speed and direction of the car as a function of time

Facts for the purpose of determining the cost of collision:

- A car is a heavy solid object that can not easily be moved.
- Colliding with a car can cause serious damage and/or injury.

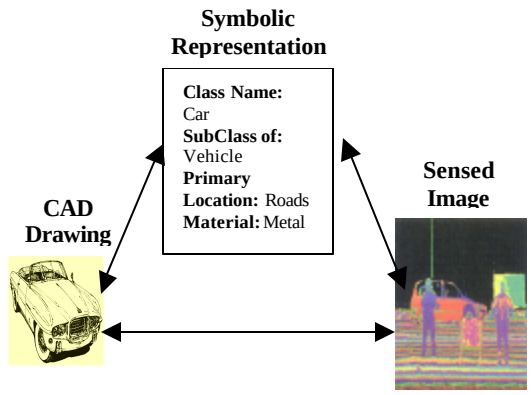

Figure 1: Different Types of Representations

Providing and using this information could only be possible if there were a mechanism to accurate associated map-based images with symbolic entities, thus provide object recognition. This would assume that the system would be able to perform the following types of steps:

1. Identify that an object exists based on various types of sensed data.
2. Identify the components of that object. This would most likely come from the combination of sensed data. For example, LADAR sensors could tell you the 3D coordinates of the obstacle. FLIR data could give you heat data about the obstacle, camera data could give you color data, etc. The combination of this data would give you the attributes of the obstacle to help you recognize it.
3. Classify the sensed image with respect to predefined classes of objects, if possible. This could be done in a number of ways, all based upon the attributes and salient features of the sensed object. If the object is stationary and is part of the terrain, one could use a priori maps and compare areas on the map with objects that are sensed to identify an object (solely by location). If prototype images (see Figure 1) and scale-invariant salient features of the various types of objects have been previously stored, one could compare LADAR images of the object with those previously stored features and images to determine if there is a match. This is similar to the MASCOT work currently being performed at the University of

Illinois at Chicago [Arlt00]. This is sometimes referred to as "Query by Example." In addition to the dimensions of the object, if cameras were used, one could compare information such as color to see if sensed objects match possible colors for objects previously stored. In addition, one could use other facts which have been stored about objects, such as the fact that car are primarily found on road surfaces.

Object recognition is a very difficult area, and a lot of research has been performed in this area with limited success. However, using a bounded case and constrained environments as a starting point, possibly comparing this information with previously obtained ground truth, a proof of concept demonstration could be performed and then generalized to a more real-life environment.

Representing Motion

For the purpose of representing moving objects, the following two assumptions are made:

- All objects are rigid masses, and all components of that rigid mass move uniformly, such that if point A on that mass moves at a given velocity and direction, all points on that mass move at the same velocity and direction. Though this may not be the case with all moving object (e.g., when a person walks, their legs and arms sway, thus moving a different speeds than the rest of their body), it should be accurate enough to simplify computation.
- All moving objects of interest touch the ground. Since the primary moving objects of interest are other vehicles, people, and animals, this should be a valid assumption. Obviously moving objects such as aircraft and flying projectiles would not be included, though for the purpose of autonomous driving, they are not of significant importance.

Recognizing and classifying objects and identifying that an object is moving are essential steps to be able to identify and represent moving objects, but it is only half the game. This is primarily because almost every object moves: trees and grass sway in the wind, a traffic light shakes back and forth due to wind and the wake of passing traffic, etc. However, these are not the types of motion that a planner cares about when determining the path a vehicle should follow and this type of information does not need to be represented in a world model. However, motion of cars, pedestrians, animals, etc. are very important to represent and track when determining the appropriate path for a vehicle to take. So how does one determine what to capture and what not to? By recognizing objects, one can query a knowledge base to determine if this is an object of interest or not within a given context. If it is, information about the movement of this object should be represented and tracked within the knowledge base while other types of movement can be ignored.

Therefore, in addition to the steps mentioned in Section 2 above, the following additional steps need to be performed to sense and track the movement of an object:

1. Identify that the object is moving
2. Correlate the "parts" of the moving object at time = t_n with the "parts" of the moving object at time = t_0.
3. Based on a single, predetermined point on the object, or on an easily identified set of salient features, track that feature or point through a series of images at known time intervals
4. Based on algorithms such as curve fitting, compute the equations corresponding to the moving object's direction, velocity and acceleration. In the case of a moving vehicle, a great deal of work has already been performed on the autonomous vehicle in computing the curves representing the possible directions the vehicle can go as a function of its wheel orientation and current velocity. These curves may also be able to be applied to sensed vehicles, and these curves could be overlaid on sensed positions of the sensed vehicle to determine which best represents its motion. The equations representing these curves could then be used to predict where the vehicle will be in the future based on its current velocity and path. In addition, one could assume that, in general, that a car primarily.

4a. There is currently considerable work ongoing in the area of lane following. In the simple case (which accounts for perhaps 80%-90% of the time) one could assume that the vehicle approximately follows the equation that accounts for the curvature of the road, and therefore one may be able to "predict" that the vehicle will continue following that curvature equation until sensed locations prove otherwise.

5. For the purpose of collision avoidance, compute the moving object's position at given time points in the future based on the identified equations.
6. Based on newly sensed data, repeat numbers 3 through 5 to refine the equations and recomputed the moving objects position at future time points.

The actual representation of the motion would then be a set of equations, in which given a time t, a corresponding position could be determined assuming the vehicle continues to move at its current speed and direction.

Symbolic Representation of Objects

In Jim Albus's book "Engineering of Mind" [Albus01], he describes a hierarchy of entities (pixels, lists, surfaces, objects, and groups), where each higher-level entity is a classification and grouping of lower level entities (see Figure 2). In this section, we expand upon the object level of the hierarchy, and introduce a simple ontology within this level for classification of objects.

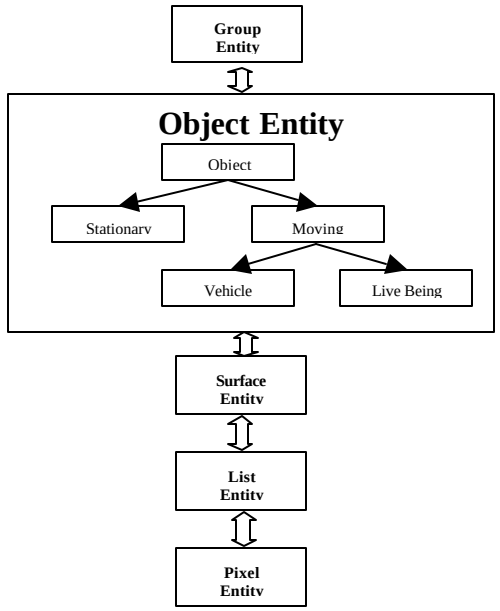

Figure 2: Expanded Object Entity in the Hierarchical Groupings of Entities

Many objects that could be encountered share similar characteristics. For example, a car and a motorcycle both are vehicles that are primarily found on the road and move at speeds between zero and 120 km/hr. It would be redundant to represent the same information twice for these two different types of vehicles. A much more efficient type of representation would be an ontology of obstacles which allow for the inheritance of attributes or characteristics from the object's parent lineage. A simple example of such an ontology is shown below, with a subset of attributes included for some classes (objects). Classes are in bold, indentation shows superclass-subclass relationships, and attributes are in italic. Attribute inheritance is supported, such that all children of a superclass inherit all of its attributes. In this case, the vehicle class would not only include its own attributes, but also all of the attributes of the **moving object** class and the **object** class. A more detailed ontology could include subclasses with multiple superclasses, many different types of relationships between classes, and a more formal type of representation.

Stuff
- **Actions**
- **Objects**
  - *Name*
  - *Position*
  - *Orientation*
  - *Color attribute(s)*
  - *Link to CAD drawing(s)*
  - *Pointer to areas in sensor representations (voxels in LADAR representation, pixels in camera representations, etc.)*
  - *Pointer to list entities which comprise the object*
  - *Pointer to surface entities comprising the object*
  - *Pointer to group entities aggregating objects*
  - *Pointer to parent class(es)*
  - *Pointer to child class(es)*
  - *Context(s) in which the object is usually seen*
  - *Rules pertaining to the object*
  - *Relationships to other objects*
  - *How bad is it if I collide with this object (this may be based on the context)*
- **Stationary Object**
  - **Road**
  - **Traffic Light**
  - **Sign**
    - **Speed Limit Sign**
    - **Stop Sign**
  - **Toll Booth**
  - **Guard Gate**
  - **Construction Cone**
  - **Road Markings**
  - **Street Lights**
- **Moving Object**
  - *Set of sensed locations (x,y coordinates) as a function of time*
  - *Velocity (both speed and direction)*
  - *Maximum velocity in different directions*
  - *Average velocity range in different directions*
  - *Acceleration (both rate and direction)*
  - *Maximum acceleration in different directions*
  - *Average acceleration range in different directions*
  - **Vehicle**
    - *Pointer to equations showing possible vehicle speed and direction given a vehicle's current velocity and wheel orientation*
    - **Car**
    - **Van**
    - **Motorcycle**
    - **Truck**
    - **Bus**
  - **Living Entity**
    - **Animal**
      - **Person**

Symbolic Representation Applied to Sensor Integration and Registration

On a tangent to what was stated above, symbolic representations could also assist in integrating "images" from disparate sensors. During the process of recognizing and classifying sensed images, one is associating a representation in the symbolic domain with a representation in the iconic domain (e.g., associating sets of pixels in the camera image with a class in the ontology). If this association is done with multiple different sensed images from different sensors, then one can correlate and register the two different images together using the common sensed image as a reference point.

Not surprisingly, one of the primary purposes of certain types of symbolic representations is for integration of different systems that use different underlying representations. Although this is usually done in the purely symbolic regime (where the external systems represent their information symbolically), this type of information integration can also be done to integrate iconic information, by performing object recognition and mapping the iconic structures (e.g., pixels, voxels) to the appropriate symbolically-represented classes.

Introduction of Driving Behaviors

To take the above discussion one step further, for the purpose of predicting where a moving object (specifically, a vehicle) will be at some time in the future, there are at least three types of information that a planner should know:

1) The acceleration, velocity, and direction of the object at the current time
2) The capabilities of the vehicle, namely, what type of motion is it capable of
3) Driving rules that the vehicle should follow (a vehicle usually stops at a red light, vehicles usually drive slower in hazardous conditions, etc.)

The first two items are addressed in the previous sections. The third item, driving rules, has been well documented in a book entitled "Driver Education Task Analysis. Volume 1. Task Descriptions" written by the Human Resources Research Organization [McKnight70]. A part of this book deals with a set of driving "situations", and the steps that a driver should perform in response to those situations. An interest exercise would be to represent a subset of those situations and behaviors in a formal matter within a knowledge base, linked to an ontology, to be able to deduce the action that an external vehicle should take in given situations. This would allow a planner to better predict the actions, and hence location, of external vehicles, and make more informed decisions as to paths to take to avoid collision.

Exchanging Position Information Among Mobile Agents via the Semantic Web

The work described in this paper is a vital component to realizing the vision of the semantic web, specifically in the areas of: sharing dynamic vehicle information among agents communicating over the Internet, for publication of object motion histories, such as for simulation evaluation, and for planning both real and simulated environments involving moving objects. Next generation traffic control systems will need to post and monitor structured information about vehicles and transportation infrastructure in order to regulate traffic flow. The military needs to post information for dynamic and retrospective analysis of simulated and in-theater operations. All these applications can benefit from structured and standard presentation utilizing the rapidly emerging infrastructure of the semantic web

The text below discusses these concepts in the context of a scenario focusing on the tracking of friendly and enemy vehicles in a hostile environment. First some background information is given, and then the scenario is presented.

As discussed above, an autonomous vehicle must constantly track the position and movements of all external obstacles for the purpose of path planning and obstacle avoidance. When little information is known about the external object, various steps need to be taken, such as recognizing what the object is and generating equations that represent the movement of the object (see previous sections). However, there may be cases in which more information is known about the external objects. For example, in times of battle, the autonomous vehicle (Vehicle A) may need to track other friendly vehicles in its section, to allow for executing section-level plans and to possible maintain a safe distance between them. In this case, much of information is already known about the other vehicles in the section, such as what the "object" is, where it is located, and the direction and velocity in which it is moving.

In addition, one of the friendly vehicles in the section may have detected a hostile vehicle that other vehicles in the section have not. In this case, this friendly vehicle would need to transmit the location and velocity of the hostile vehicle over the semantic web to other friendly vehicles in the section. What is needed is a way to accurately communicate this information.

This communication must happen in a fashion in which the semantics of the message are completely and

unambiguously understood. The semantic web could indeed provide this forum, but only if it is based upon a well-defined ontology consisting of types of obstacles and objects in which the friendly vehicles could encounter, as well as very well-defined characteristics of those objects.

The simple ontology provided in the previous section provides a starting point for introducing terminology and related characteristics to aid in the communication among mobile agents via the semantic web. It provides a classification of the types of objects that one would expect to sense in the environment, along with the appropriate types of characteristics for those objects. Position and velocity are some of the characteristics that one would wish to know about a moving obstacle.

Imagine the situation where a friendly vehicle (Vehicle A) senses a hostile tank and begins to track its position and velocity. It would immediately need to notify other friendly vehicles in its section via the semantic web to make them aware of the enemy tank. Unfortunately, the sensors on Vehicle A have a very limited field of view, so the enemy tank can only be tracked by Vehicle A for a short amount of time before it leaves its field of view. However, Vehicle A was successful in capturing the movement of the enemy tank, and was able to accurately predict where the enemy tank would be at a given time in the future. This information is sent over the semantic web to another friendly vehicle (Vehicle B) in the section whose field of view encompassed the location where the enemy tank was expected to be, and Vehicle B then senses the enemy tank and tracks its motion. This continues from vehicle to vehicle, until the enemy tank is sufficiently out of range, or is considered an imminent threat and is therefore fired upon.

Since the enemy vehicle was identified as being a tank, various characteristics of a tank are known a priori, such as the tank's maximum speed in both the forward and reverse direction, as well as its ability to turn in tight spirals. This information allows the system to more accurately predict not only where the enemy tank will be at a point in the future, but to confidently predict where the tank can not be.

Even with this, the problem is not solved. The individual vehicles need to be able to exchange position and velocity information in an unambiguous way. Having common terminology and semantics, such as the simple ontology presented in the previous section, helps to accomplish this. Initial work has just begun to integrate this ontology with the current SUO [Pease02] work, thus providing more formal semantics for the concepts introduced.

Even with this, the problem is not solved. Each sensor on the vehicle is egocentric (i.e., the vehicle doing the sensing is perceived as the center of the universe) and all sensed objects' positions are represented in a coordinate frame in which the autonomous vehicle is at the origin. Thus, not only does the position and velocity information have to be passed (and unambiguously understood) among vehicles, but so does a coordinate transform that can map a sensed object from the perspective of one vehicle's coordinate frame to the coordinate frame of another vehicle. Thus the position of any friendly vehicle would need to be known and maintained with respect to the position of any other friendly vehicle in the section.

Not only would this position information need to be shared among the friendly vehicles in the section, but it would also most likely need to be shared with some type of central command that would continuously monitor enemy vehicles such as to develop a section-level plan to combat the enemy vehicles, if necessary. Thus, this position information would also need to be transformed into a central coordinate frame, for the purpose of tracking outside of any individual vehicle, and sent to the central command via the semantic web.

Conclusion

This paper describes a lot of challenges and research areas related to the areas of object detection and symbolic representations, while only skimming the surface on providing concrete solutions on how to resolve these challenges. The paper represents the very beginning of a research effort that is getting underway at NIST, which is exploring the incorporation of ontologies and other types of symbolic representations into a control hierarchy to control an autonomous vehicle.

It is impossible to separate the representation of an obstacle with the process in which it is recognized, which is why a good portion of this paper deals with the steps that must occur during the object recognition process. Many efforts in the past have attempted to develop representations (ontologies) of objects and actions without any clear indication of how they were expected to be used. Although there is value at doing this at the highest levels of ontology building (with the expectation of having these general ontologies be specialized for domain-specific purposes), any attempt to develop a domain ontology needs to have a clear understanding of not only what types of concepts need to be represented, but what specifically needs to be represented about those concepts. In the case of representing sensed obstacles for autonomous driving, one would also need to know how those objects were perceived, so the appropriate characteristics of the object could be included in the representation (e.g., salient features, CAD drawings, etc.).

In this paper, we discussed the importance of recognizing and representing both stationary and moving obstacles, as well as the value that one would gain by linking these representations to an ontology of obstacles. This ontology could not only contain a taxonomy, with attributes, of

obstacles, but also a set of rules and statements about the obstacles. These rules and statements could facilitate operations such as obstacle detection, location prediction, and associating a cost/risk of colliding with the obstacle. For example, if the rules of driving behaviors were embedded in the knowledge base, one could more accurately predict where an external vehicle could be at a time point in the future, given a set of driving conditions, to help better avoid colliding with that object.

References

[Albus01] Albus, James S., Meystel, Alexander M. eds. 2001. *Engineering of Mind: An Introduction to the Science of Intelligent Systems* New York: John Wiley and Sons.

[Arlt00] Arlt B., Brause R., Tratar E., *MASCOT: A Mechanism for Attention-based Scale-invariant Object Recognition in Image Databases*, Internal Report, Computer Science Dept. (FB20), J. W. Goethe University Frankfurt/Main, 2000

[Hong02] Hong T.H., Balakirsky S., Messina E., Shneier M., "A Hierarchical World Model for an Autonomous Scout Vehicle", *Proceedings of the SPIE AeroSense Conference Vol. 4715*, Orlando, Florida, April 1-5 2002.

[McKnight70] McKnight, James and Adams, Bert, *Driver Education Task Analysis. Volume 1. Task Descriptions*, Human Resource Research Organization, Department of Transportation, National Highway Safety Bureau, 1970.

[Pease02] Pease, A., Niles, I., IEEE Standard Upper Ontology: A Progress Report, Knowledge Engineering Review, Special Issue on Ontologies and Agents, Vol 17, 2002.

[Shoemaker99] Shoemaker C., Bornstein J., Myers S., Brendle B., "Demo III: Department of Defense testbed for unmanned ground mobility", *Proceedings of the SPIE Col. 3693, AeroSense Session on Unmanned Ground Vehicle Technology*, Orlando, FL, April 7-8, 1999.