FUSION OF GPS AND MACHINE VISION FOR ABSOLUTE VEHICLE POSITIONING

Andrew Rae, PhD Candidate
Otman Basir, Associate Professor
Pattern Analysis and Machine Intelligence (PAMI) Research Group
Department of Electrical and Computer Engineering
University of Waterloo
Waterloo, ON N2L 3G1
Canada
rarae@uwaterloo.ca
obasir@uwaterloo.ca

ABSTRACT

This paper describes the use of data from a Geographical Information System (GIS) as *a priori* knowledge to be exploited for reliable and consistent vehicle positioning. The line-of-sight approach taken by GPS creates location inaccuracies especially in urban areas, or leads to complete signal loss when underground, in tunnels, etc. Many attempts have been made to complement GPS with dead reckoning sensors for the purpose of more reliable vehicle positioning; however, such sensors typically suffer from accumulated error. Our approach incorporates a vision system to detect road features and landmarks that can be matched to data from a GIS to determine the position of the vehicle. This information can be fused with the GPS position estimate to increase the reliability of the overall position estimate. The GIS knowledge base incorporates several types of information on different layers, such as road shape, location of landmarks such as buildings, and surrounding terrain. Furthermore, the information available from GIS about the environment surrounding the vehicle position allows constraints to be placed on the possible vehicle positions, creating a positioning system that is highly reliable and aware of its environment. An issue with the type of GIS data we have used is the disparity in representation of features versus a visual representation, which we shall discuss herein and provide suggestions on how to address.

INTRODUCTION

The advent of satellite positioning systems such as the Global Positioning System (GPS) has made it possible to accurately determine one's location on Earth. However, the line-of-sight approach taken by GPS creates location inaccuracies especially in urban areas (a problem referred to as "multipath") or leads to complete signal loss when underground, in tunnels, etc. Many attempts have been made to complement GPS for the purpose of more reliable vehicle localization with dead reckoning sensory data, e.g., odometry and inertial sensors. However, these sensors suffer from accumulated error, particularly the inertial sensors which require integration. Our approach follows the notion that visual information and a spatial reference can be used to determine position. This paper describes the use of Geographic Information Systems (GIS) data as this spatial reference, against which visual feature observations are compared to determine vehicle position. In particular, we are considering visual features which vary continuously as a function of vehicle position, e.g., road curvature and distance to objects in the environment. By the further restriction that measurement errors must be Gaussian, we are justified in using a Kalman filtering method for the purpose of estimating vehicle position over time (Aralumpalam, 2002).

It is our intention to use vision for the purpose of "longitudinal positioning" (vehicle position along the length of the road) versus "lateral positioning" (vehicle position relative to the boundaries of the current lane) (Laneurit, 2003). This method requires *a priori* knowledge of the location of landmarks and the value of features in the environment surrounding the vehicle. This approach to positioning is beneficial because it establishes a direct link to the vehicle environment by observing distinct features and using them to produce a position estimate. This is in contrast with traditional satellite positioning methods or dead reckoning methods and should help to alleviate some of the problems of these respective approaches. Specifically, we expect that this method will help to compensate for multipath and signal loss in GPS position fixes, and to remove accumulated error issues in dead reckoning.

The remainder of the paper is organized as follows. First, the proposed method based on Kalman filtering and feature matching is briefly discussed. Second, the use of GIS data to provide *a priori* knowledge of the environment is discussed, in particular what we believe to be advantages to its use as well as the issue of "visual disparity" (disparity between features observed visually and their representation in GIS data). Third, we provide some simple simulation results that illustrate the benefit of using visual information in vehicle positioning, as well as the importance of reducing/eliminating visual disparity. Finally, we conclude with some future work and concerns, such as suggestions on how to handle the disparity issue and make visual information viable for positioning.

PROPOSED METHOD

It can be argued that visual information is used to determine position in a bottom-up manner (Figure 1). At the bottom levels, features of the environment must be detected and recognized. At an intermediate level, the recognized features can be used to determine an approximate position. For example, by recognizing a specific cluster of buildings or other landmarks, one can determine the road on which they are traveling. At the highest level, position can be determined with greater precision as the most likely position from which the recognized features are being viewed. To continue the previous example, this could involve estimating the distance between the vehicle and the recognized buildings or landmarks.

Implicitly contained within this bottom-up framework is the notion that visual features are associated with positions in some way. There needs to be a knowledge base of some kind which relates detected visual features to positions. One can consider this knowledge base as the system's internal model of the world, against which visual information is matched in order to determine the appropriate position from which it is being viewed.

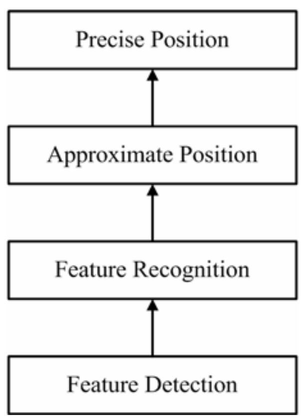

Figure 1. A bottom-up model of visual position estimation.

By assuming we have access to such a model of the world, we have developed a method to estimate vehicle position by complementing position estimates attained from GPS with visual information. Visual information essentially places constraint on the possible locations for the GPS receiver by comparing observed visual features with features derived from the world model. The latter being associated with positions in the world allows this comparison to influence the final position estimate. This method thus focuses on the highest level in Figure 1, and assumes that lower level detection and recognition stages are taken care of.

Details of this positioning method, which is in effect the fusion of two complementary data sources, are reported previously in (Rae, 2007). Our method uses a Kalman filtering scheme to combine vision and GPS data and produce a vehicle position estimate. This is a common approach taken by other researchers (e.g., Caron, 2006; Chausse, 2005; Laneurit, 2003; Wang, 2005) in attempting to make GPS more robust to issues such as multipath and signal loss. The Kalman filter (KF) is a popular method producing statistically optimal state estimates for linear systems affected by additive Gaussian noise. These linearity and Gaussianity requirements can be highly restrictive; as such, researchers have developed non-linear and non-Gaussian approximations possessing many of the same computational benefits as the standard KF but which produce suboptimal results (Aralumpalam, 2002). KF approaches however share the common requirement that variables must be continuously valued. In our case this places further restrictions on the type of features that can be used for positioning; features must depend continuously on vehicle position, versus features which take on a discrete value for a series of positions. Examples of features which depend continuously on position include road curvature and distance to static objects.

GEOGRAPHIC INFORMATION SYSTEM (GIS) WORLD MODEL

We propose to use GIS data as the world model for our system. GIS data provides spatially referenced information about the features in the local area, typically in a layered format where each layer corresponds to a different class of information. For instance, roads, buildings, forested areas and terrain elevation would all be stored as separate layers of information. The goal is to match the features contained in the GIS data, which are associated with positions, with the features observed using vision.

Some past approaches to vehicle positioning and localization have operated along this same principle of feature matching (e.g., Georgiev, 2002; Jabbour, 2006). The crucial difference between these approaches and our own lies in the GIS world model; while some previous methods have used GIS, they tend to focus on the need for highly accurate mappings of primitive visual features, such as distinct corners and lines. Such approaches require extensive training and data collection to create highly detailed world models which contain the set of features viewed from each possible location. This approach is commonly used in the Simultaneous Localization and Mapping (SLAM) problem of robotics and represents the continuation of a trend better suited for indoor environments, which are typically smaller and less complex than outdoor environments. An outdoor map composed of visual primitives would likely require a large amount of storage, be prohibitively time-consuming (and therefore expensive) to update, and in all likelihood be dependent on the vision system used to create it.

Our approach seeks to use existing GIS information describing the location and shape of objects in the world, rather than construct a custom layer. This is also how we as humans position ourselves visually; we recognize known landmarks such as buildings, rather than a collection of distinct corners and lines (although these may help us to recognize our landmarks). There are a number of benefits for such an approach:

1. The data is not system dependent and can therefore be considered "objective."
2. Many such GIS databases are already compiled which contain features useful for vehicle positioning (e.g., Figure 2).
3. Increased flexibility since many different types of features can be used provided they vary continuously with position.

The latter point is important for the robustness of the approach in terms of operating in different environments, and if one were interested in maximizing positioning ability by using only features which are most relevant for the task.

When matching features obtained by a vision system to features as they appear in a GIS or other type of map of the environment, one must deal with the problem of disparity, which is caused by a difference in the way information is presented in both media. A vehicular vision system is constrained by the placement of the camera(s), which must be mounted on the vehicle and therefore views the road and surrounding environment from a nearly horizontal angle. Therefore the camera perspective as well as effects such as lens distortion and pixel resolution, affect the appearance of features in captured images. The information in a typical GIS layer would be free of such distortion and perspective issues. The disparity in representation, however, can create problems when matching features from both sources (Manyika, 1994).

An advantage for systems which use collections of visual primitives as features, such as lines or corners, is that the world model is typically "built" using the same system in a training phase and hence the problem of disparity is resolved. The approach we take in reducing disparity between visual features and features from pre-existing GIS data is to view the vision system as a transformation from real-world features (as contained in our GIS data) to visual features. Approximating such a transformation and applying it to GIS features would reduce the disparity issue commensurate with the quality of the approximation.

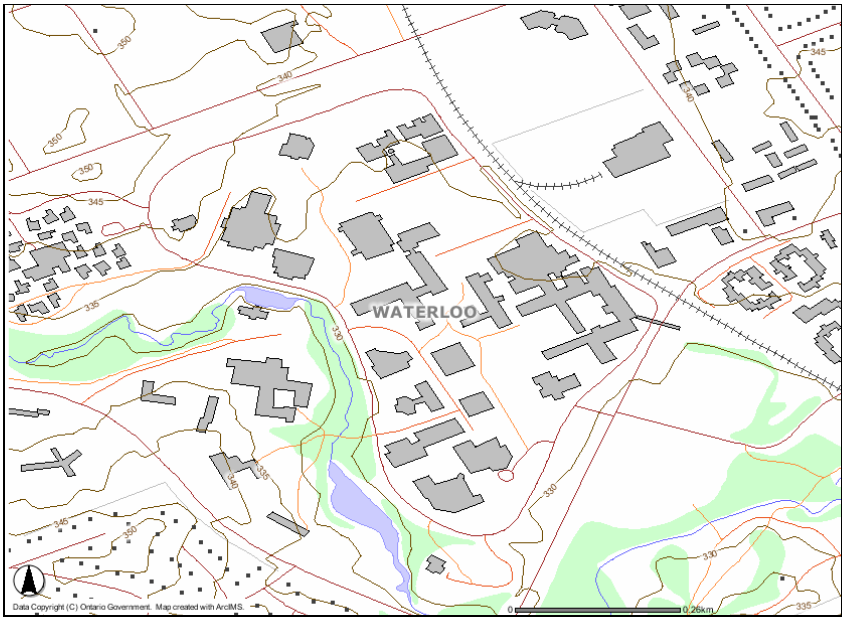

Figure 2. Illustration of many layers of data useful for positioning in the area surrounding the University of Waterloo in Waterloo, Ontario, Canada. Pertinent information includes forested areas (green polygons), water (blue polygons), building footprints (grey polygons), roads (maroon lines) and elevations (black lines). (Source: Ontario Basic Mapping (OBM), www.geographynetwork.ca).

Consider as an example the visual estimation of road curvature ahead of the vehicle. In itself road curvature is an important and useful feature for many applications in ITS. Examples include lane keeping, lane departure warning and vehicle rollover warning. It is often measured as a by-product of visual lane boundary detection (e,g,, Chapuis, 2002). Many of these detection schemes take advantage of the fact that roads, particularly highways, are built under a constraint of slowly changing curvature. This has led to two predominant approaches to lane boundary detection: the first assuming the section of road in the image has constant curvature, the second that the curvature varies linearly with arc length along the road (this is known as a clothoid model). One could model the former approach by selecting a single value to represent the curvature over a specified length of road ahead of the vehicle, say the mean curvature.

Illustrated in Figure 3 is an example of the type of disparity that could arise in feature matching. Consider the values of actual and observed road curvature for the indicated arc length on the corresponding road segment; the road curvature is at its maximum positive value (the road is bending to the right), while the curvature measured by a vision system mounted to a vehicle at that position is a negative value (since the road straightens out and bends to the left ahead in the section of road immediately ahead of the vehicle). The disparity in this situation would produce large feature matching errors which affect the ability to determine position visually. There is therefore a need to learn the transformation from true features to visual features for each feature extraction algorithm used, thus representing a training phase for the system.

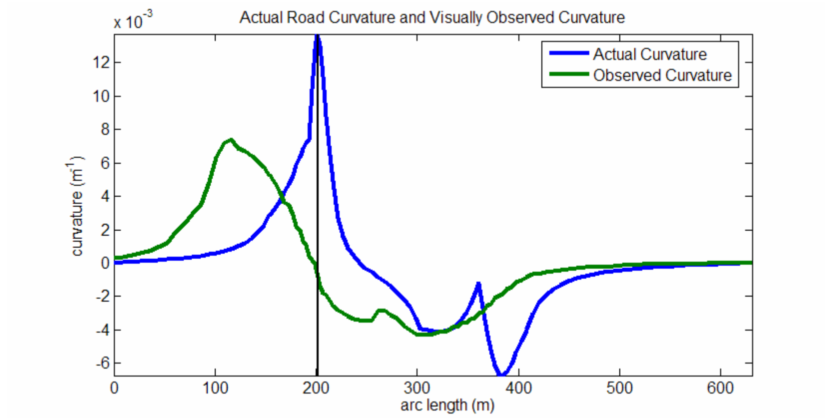

Figure 3. Disparity in road curvature estimates; actual curvature (blue) and visual observation (green).

SIMULATION RESULTS

Results are shown in Figure 4 to illustrate i) the improved positioning performance offered by integrating GPS position estimates with visual feature data, and ii) the need to reduce disparity between visual features and GIS features. The vehicle was traveling on the road for which the curvature is displayed in Figure 3. At 1s intervals, a position estimate from GPS was acquired and a visual measurement of road curvature was made. Visual road curvature measurement was simulated by adding random noise to the value of the visual observation model corresponding to the actual vehicle position at each 1s interval. GPS position estimates were simulated using reported error models for a Coarse Acquisition, Standard Correlator-Width receiver (Rankin, 1994).

The results show that when the true road curvature is matched to visual curvature measurements, the resulting position error is higher than standalone GPS. Conversely, when the visual observation model is matched to curvature measurements, position error is significantly improved versus standalone GPS due to the reduction in disparity.

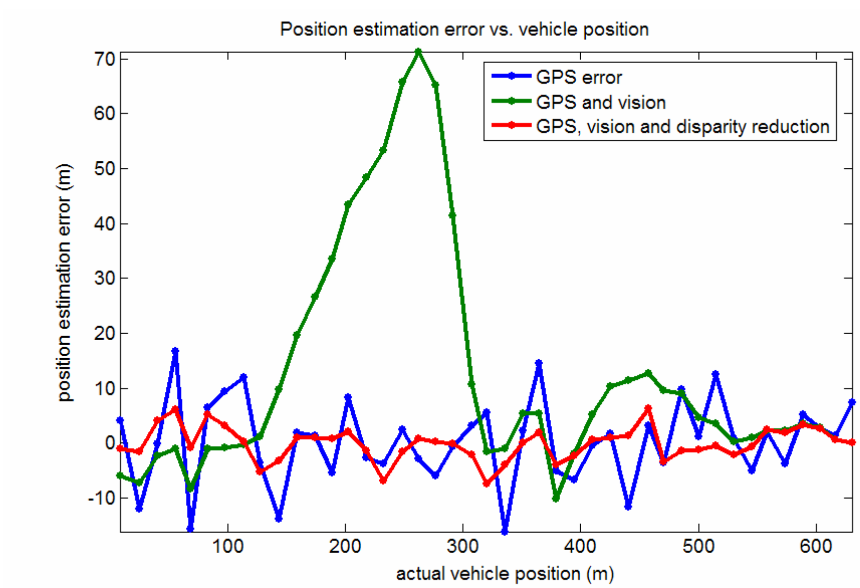

Figure 4. Vehicle positioning errors; the use of visual features without accounting for disparity (green) produces worse estimates than standalone GPS (blue), while accounting for the disparity (red) improves on GPS estimates.

These results indicate that disparity reduction is necessary for positioning accuracy to be improved. In this paper we have presented a hypothetical situation in which the transformation from real-world features (contained in the GIS) to visual feature observations is known exactly. In reality, one cannot expect this to be true. It may thus be necessary to use neural networks and other machine learning methods to determine analytically the form of such transformations.

CONCLUSIONS

We discuss in this paper the use of GIS data in a system that fuses visual measurements of real-world features with position estimates from GPS for the purpose of vehicle positioning. The use of vision in this manner is an important contribution to vehicle positioning since it creates a link to the surrounding environment of the vehicle, unlike dead-reckoning or GPS technologies.

The GIS data provides a world model for our system, from which features are derived for the purpose of matching with visual feature measurements. We have shown in this paper that the manner in which these GIS-derived features are used is particularly important; any disparity that exists between how features are presented in the GIS data and how features are "seen" by the vision system must be reduced/eliminated in order for our approach to be successful. In a general sense, we intend in the future to use machine learning techniques such as neural networks to learn the "disparity transformation" for each visual feature.

The success of this positioning method also depends on the quality of the GIS data, namely its accuracy and currency. Data that is inaccurate will translate into inaccurate position estimates since a disparity exists between the real-world and the model provided by GIS. Similarly, data that is out-of-date will create problems in situations where features are detected visually but are not present in GIS, say when new buildings have been erected or new roads built since the GIS database was compiled. A registration process may thus be required, to determine which objects observed in a scene are present in GIS and which are not.

REFERENCES

Aralumpalam, M. S., S. Maskell, N. Gordon, and T. Clapp (2002). A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking. *IEEE Transactions on Signal Processing*, *50*(2): 174-188.

Caron, F., E. Duflos, D. Pomorski and P. Vanheeghe (2006). GPS/IMU data fusion using multisensor Kalman filtering: introduction of contextual aspects. *Information Fusion*, *7*: 221-230.

Chapuis, R., R. Aufrere and F. Chausse (2002). Accurate Road Following and Reconstruction by Computer Vision. *IEEE Trans. ITS*, *3*(4): 261-270.

Chausse, F., J. Laneurit and R. Chapuis (2005). Vehicle localization on a digital map using particles filtering. *Proceedings of the 2005 IEEE Intelligent Vehicles Sumposium*: 243-248.

Georgiev, A., and P. K. Allen (2002). Vision for Mobile Robot Localization in Urban Environments. *Proceedings of the 2002 IEEE/RSJ International Conference on Intelligent Robots and Systems*: 472-477.

Jabbour, M., P. Bonnifait, and V. Cherfaoui (2006). Enhanced Local Maps in a GIS for a Precise Localization in Urban Areas. *Proceedings of the 2006 Intelligent Transformation Systems Conference*: 468-473.

Laneurit, J., C. Blanc, R. Chapuis and L. Trassoudaine (2003). Multisensorial data fusion for global vehicle and obstacles absolute positioning. *Proceedings of the 2003 IEEE Intelligent Vehicles Sumposium*: 138-143.

Manyika, J. and H. Durrant-Whyte, (1994). *Data Fusion and Sensor Management: A Decentralized Information-Theoretic Approach*, Ellis Horwood, New York, chapter 6.

Rae, A. and O. Basir (2007) A Framework for Visual Position Estimation for Motor Vehicles. *Proceedings of the 2007 Workshop on Positioning, Navigation and Communication*: to appear.

Rankin, J. (1994). GPS and Differential GPS: An Error Model for Sensor Simulation. *Proceedings of the 1994 IEEE Position, Location and Navigation Symposium*: 260-266.

Wang, J., S. Schroedl, K. Mezger, R. Ortloff, A. Joos and T. Passegger (2005). Lane Keeping Based on Location Technology. *IEEE Transactions on Intelligent Transportation Systems, 6*(3): 351-356.