A Context-Based Approach to Vehicle Behavior Prediction

Stewart Worrall, Gabriel Agamennoni, Juan Nieto, and Eduardo Nebot
Australian Centre for Field Robotics, University of Sydney

Abstract—Despite the best efforts of research and development carried out in the automotive industry, accidents continue to occur resulting in many deaths and injuries each year. It has been shown that the vast majority of accidents occur as a result (at least in part) of human error. This paper introduces the model for the Intelligent Systems for Risk Assessment (ISRA) project which has the goal of eliminating accidents by detecting risk, alerting the operators when appropriate, and ultimately removing some control of the vehicle from the operator when the risk is deemed unacceptable. The underlying premise is that vehicle dynamic information without contextual information is insufficient to understand the situation well enough to enable the analysis of risk. This paper defines the contextual information required to analyze the situation and shows how location context information can be derived using collected vehicle data. The process to infer high level vehicle state information using context information is also presented. The experimental results demonstrate the context based inference process using data collected from a fleet of mining vehicles during normal operation. The systems developed for the mining industry can later be extended to include more complex traffic scenarios that exist in the domain of ITS.

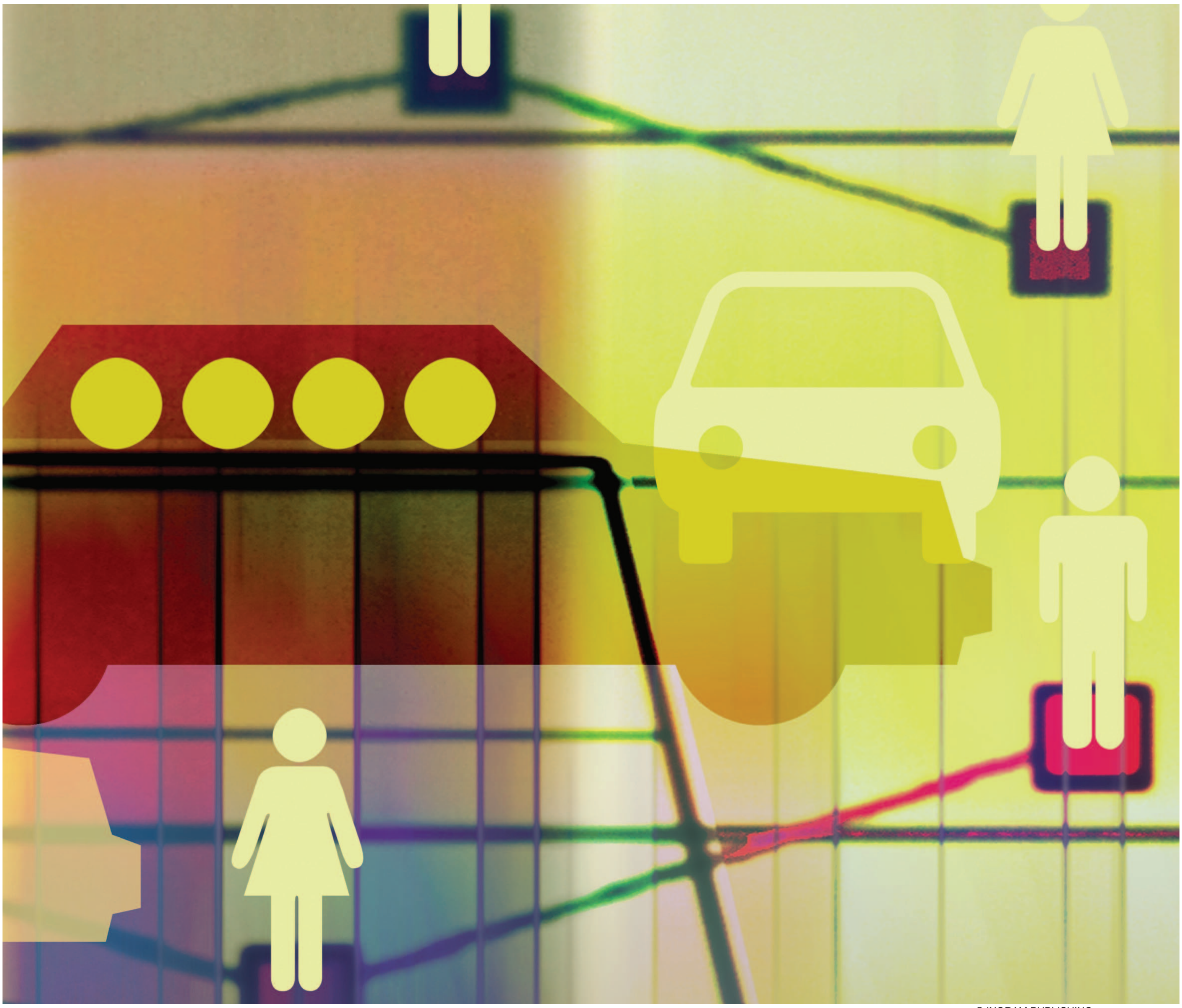
I. Introduction

One of the primary goals of implementing technology in the Intelligent Transportation Systems (ITS) domain is to improve safety by reducing risk, with the intended consequence of minimizing accidents. The short term goal of this endeavour is to detect high risk situations and alert the operator of the vehicle in an appropriate manner [1]. The ultimate goal is to eliminate the risk by removing the human factor, which is a contributing factor in the majority of vehicle accidents [2]. The human contribution to accidents can be reduced or even eliminated by removing some of the control of the vehicle when a risk is determined to be unacceptable with a high degree of certainty.

Digital Object Identifier 10.1109/MITS.2012.2203230
Date of publication: 25 July 2012

There are several difficulties that make determining safety risk a non-trivial problem. Determining what constitutes a "high risk" situations requires a high level understanding of the context of the situation. Context is required to encapsulate higher level information (meaning) to the vehicle state in an effort to improve the determination of risk, and minimise false alarms. A "false alarm" is an instance where an operator is unnecessarily warned or an action is taken to a risk that is not real. We must consider that false alarms pose a considerable problem as they have the potential to distract, annoy, provide incorrect instruction, provide an inappropriate action, etc. To reduce false alarms, it is necessary to consider whether the state of the vehicle relative to the other vehicles and the environment have high risk given the current context information. The list

 1939-1390/12/$31.00©2012IEEE



below identifies the major contextual factors that can alter the perceived risk facing a vehicle.

- **Dynamic Vehicle State:** Understanding the dynamic properties of the vehicle and other vehicles/objects in the local environment is required to determine safety risk. The vehicle parameters such as speed, acceleration and heading determine the future possible motion of the vehicle. The vehicle state can also include measures of the vehicle control inputs such as the state of the accelerator and brake pedals.
- **Location Context:** The location of the vehicle is a major factor in determining whether the current vehicle state constitutes "high risk" behavior. The road location could be described as a residential road, highway, etc, and the area location could be described as a parking lot, intersection, petrol station, etc. The expectations for safe vehicle behavior changes depending on the high level understanding of this location context.
- **Environmental Context:** The environmental conditions can alter the dynamic properties of the vehicle (traction), and the ability to detect threats due to rain, dust, fog, etc.
- **Human Context:** The ability of a human to operate a vehicle depends on many complex factors including skill, fatigue, experience and many others. The determining of safety risk would be improved by being able to quantify these factors, though many are practically unobservable.

This paper presents an approach for determining location context information and inferring the vehicle dynamic properties (what the driver is currently doing), developed as part of the Intelligent Systems for Risk Assessment (ISRA) project. The ISRA project aims to develop the

> Information is passed to the context-based inference engine which determines the current vehicle behavior based on the vehicle's dynamic state, encoded expert knowledge, and location context information.

fundamental perception technology required to evaluate new safety metrics and to provide operators with complete situation awareness of the current area of operation. Figure 1 shows an overall picture of the main blocks encompassed in the project.

The blocks in Figure 1 can be divided into two main categories: real-time and offline. The real time system in the vehicles begins with on-board sensing and communication to provide each vehicle with the state information of itself, and all other nearby vehicles. This information is passed to the context based inference engine which determines the current vehicle behavior based on the vehicles dynamic state, encoded expert knowledge and location context information as illustrated in Figure 2. It is planned for the predicted vehicle behavior from the inference engine to be used to generate a probabilistic measure of risk, which triggers actions such as graphics, sound and potentially removing control of the vehicle from the operator. The risk analysis component of the system is planned for future work of the ISRA project.

The vehicle state information including position, interaction and event data is collected for off line analysis. This data is used to determine the location context information, and to enable the learning and encoding of expert knowledge through analysis and simulations. Expert knowledge can be encoded in several ways. For example, Section III-B introduces a set of feature functions that capture knowledge about context information and vehicle dynamic relationships. Rules can also be defined based on the constraints of the road rules. The data collected is also required to validate and detect failures in the hardware and software systems.

There is existing research that considers some situation context information, and uses this to alter the behavior of the vehicle and/or model of the environment [3], [4]. Other work towards situation based analysis of risk includes [5], where a goal-based approach is proposed to anticipate the motion of drivers. The authors design a model

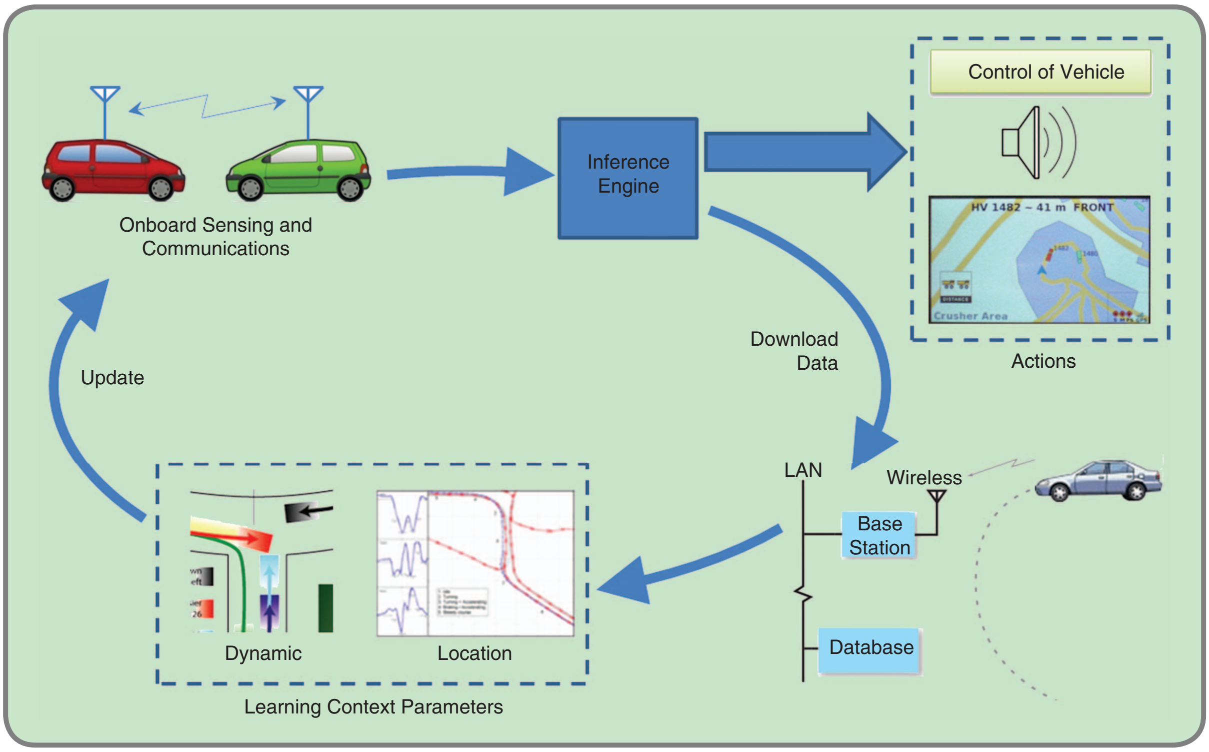


FIG 1 Overall picture of the intelligent safety system (ISRA) proposed in this project.

for predicting vehicle manoeuvres based on situation-specific motivations and test their model in a simulated highway environment. Another approach along the same lines is that in [6]. Here, a spatio-temporal situation model is proposed for inferring contextual information from semantic descriptions of the state of traffic. The model is fully probabilistic and is tested with synthetic data from a race car simulator. Another probabilistic formulation appears in [7]. Their work presents a filter for estimating the behavior of traffic participants and predict their future trajectories. Prediction is performed by first recognizing the situation class, which is itself derived from the local situational context.

To detect high risk vehicle scenarios, it is essential to have an understanding and meaning of the vehicle's location.

This paper focuses on the learning location context information, and using context information and the dynamic vehicle state to infer a high level estimate of vehicle behavior. Section II introduces the location context information, and describes techniques for building a map with sufficient detail to be used for estimating vehicle motion. A road network and areas of importance are extracted from collected vehicle data. Section III uses context based inference to determine the current high level vehicle behavior using the vehicle state and context information.

Section IV shows experimental results using data collected from a collision avoidance system installed on a fleet of mining trucks. A typical mine has very large machinery operating in often difficult environmental conditions, and there is clear motivation in the industry to reduce incidents and improve safety. The systems developed for the mining industry can later be extended and generalized to the domain of ITS. A mining environment is diverse, with rough, poorly defined roads/terrain, though the problem is constrained as the mine is a closed environment. Vehicle access is restricted and the roads of the mine, while they are highly dynamic, are bounded by the limits of the mine. Finally, conclusions are presented in Section V.

II. Location Context

A. Overview and Definitions

To detect high risk vehicle scenarios it is essential to have an understanding and meaning of the vehicle's location. The location provides meaning to other vehicle parameters such as velocity, safe travelling distances, etc. For example, a high vehicle velocity would be considered dangerous if the vehicle was approaching an intersection, or in a parking lot while it could be acceptable on a highway.

The location can be represented by a digital road map, which can be used for localization, guidance and navigation. The road map network is also important to understand the potential range of future motion of the vehicle. At intersections, the road map encapsulates the potential choices available to the vehicle operator.

A digital road map is defined as a graph where roads are the graph edges and the intersections/junctions form the graph nodes. Existing, commercially available digital maps usually consist of roads where bidirectional roads are defined as single graph edges. To store high accuracy lane information and other lane properties, we define each lane as a separate graph edge. This means that a bidirectional road will be represented by two graph edges, one for each direction of travel.

While it is possible to extract some road information using graphical techniques (from video/images), high precision lane information and velocity modelling can only be obtained by using vehicle trajectories. This is particularly the case in areas where lanes are not well marked, or on dirt roads. Vehicle data is collected as part of the ISRA system, described in Section IV.

This paper considers three main elements to construct a road map:

- **Road:** Roads are modeled as a unidirectional graph edge representing a single lane. There is an expectation that a vehicle will continue travelling the same road until they reach an intersection, or a defined context area. The motion of a vehicle on a road is bounded by a velocity model that encapsulates the maximum velocity (speed limit) and other road properties such as curvature.

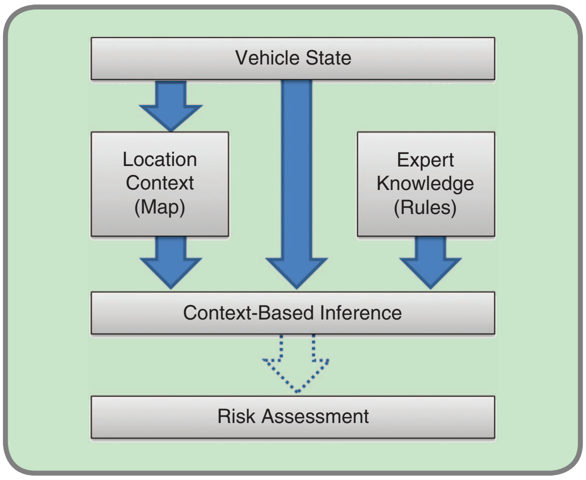


FIG 2 Schematic of the inference engine.

To generate a road map graph using vehicle trajectories, the raw data is sampled, relational links between the samples are created, and complex areas are bounded.

- **Intersection:** The junction between roads form an intersection, which is represented by a graph node. The motion of a vehicle within an intersection is well defined, but contains multiple possible scenarios based on the set of roads that leave the intersection. Intersections are particularly important for safety applications [8] as there is potential for interaction with other vehicles as the roads meet from different directions.
- **Context Area:** An area where the potential vehicle motion is not well defined is by definition an area of interest. This is defined as a "Context Area," and represents the boundary of an area representing an unstructured, or complex structured environment such as a parking lot, open field, and any other area where vehicles can move or change direction unpredictably. Context areas generally have some high level meaning that is difficult to detect without human understanding and reasoning.

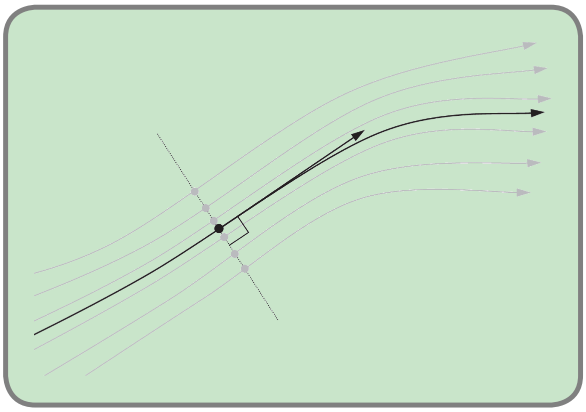


FIG 3 Finding the principle road path (centerline) of a road as the average of a set of traces.

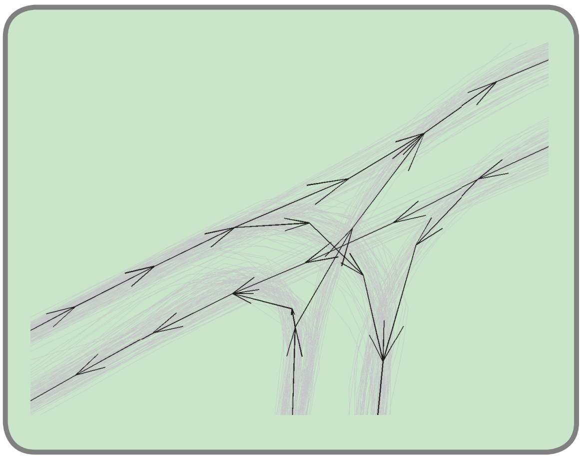


FIG 4 Linking sample points in a complex intersection.

This section describes the generation of a road map graph created using collected vehicle trajectories. The process is described in three steps: Sampling the raw data, creating relational links between the samples and bounding complex areas.

B. Sampling

The position and velocity information collected from the vehicles can be used to determine the location of the road lanes. The goal is to determine the average trajectory taken by a vehicle, which is defined here as the "Principle Road Path" (PRP) [9] as illustrated in Figure 3. The main difficulty with determining this path is that there is uncertainty in the position which can lead to ambiguity when assigning each trajectory to a corresponding path.

This problem has been approached by analyzing the position and heading information of individual data samples [10], and by considering the curves formed by the interpolation of each data point [9]. To resolve the ambiguities of associating data traces to road paths, a soft assignment can be made taking into account the uncertainty of the assignment until enough evidence is available to reduce the uncertainty [9].

Both approaches require initial estimates of the location of the PRP that can be iteratively improved as additional traces are incorporated into the model. These seeding points are generated by taking each trajectory and a decision is made whether the seed is sufficiently far (in distance and heading) from the existing points. Once generated, the sample set is iteratively improved using additional data traces.

C. Relational Linking

The topology of the road network is established by determining the relationship between the road samples defined in the preceding section. The vehicle trajectories are used to determine the transitions between the samples. Each detected transition strengthens the relationship between the samples with an increase in the count for the corresponding graph edge.

Figure 5 shows an example of a typical intersection with transition probability marked. It can be seen in this figure that there are redundant graph edges, in this case caused by lateral variation in the trajectories of vehicles turning left at the intersection. Over-sampling increases the number of samples, and consequently increases the probability of redundant edges. The graph can be optimized by removing redundant links without compromising the integrity of the road network topology. To decrease the number of redundant sub-graphs, the weight of the edges of each sub-graph is maximized [9] which effectively determines the most traversed set of edges in the graph.

In a context area, the vehicles do not display a dominant behavior pattern that can be modeled using the sampling and linking techniques.

Figure 6 shows an example of an area where the linking algorithm breaks down. In an unstructured area where the vehicles can travel in many different directions, it is not possible to determine the Principle Road Path because it is undefined. The lack of predictability of the vehicle motion makes the area by definition a high risk area when compared to the well defined motion of a road or intersection. The next section examines how this area can be bounded.

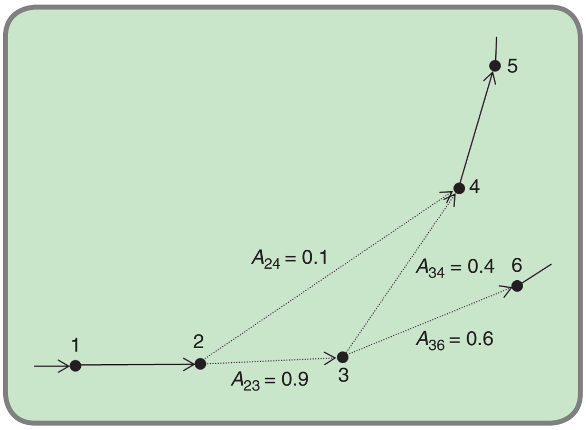


FIG 5 Finding and removing redundant edges.

D. Areas of Importance

An area of importance, defined here as a context area, is an area where the motion of the vehicle is difficult to predict. The vehicles do not display a dominant behavior pattern that can be modeled using the sampling and linking techniques outlined in this paper. Once these areas are bounded and the meaning identified, rules defining safe behavior can be defined based on the type of context area.

A sample is considered part of an important area if there are multiple links to nearby samples. This means that an intersection is also defined as an area of interest. The assumption is that the underlying shape of the important area can be constructed using the set of nearby road samples containing multiple links.

The underlying shape of a set of samples can be determined by using the alpha shape technique [11]. This technique uses a circle of radius α to essentially 'scoop' out the area surrounding the border of the points. The process begins by taking a set of points to be included in the shape (which in this case is the set of samples containing more than one link), and determining the circumcircles that can be created using each pair of points. The boundary of the shape is defined by all the circumcircles that do not contain one of the other points in the set. This means that an α-shape can contain holes and concave areas. As α approaches infinity, the shape approaches the convex hull. The value of α was selected to be twice the size of the distance between samples, meaning that all samples that should be joined are within the range of the nearest points when creating the circumcircles.

The alpha shape technique is illustrated in Figure 7. The polygon as created using the road sample points is too tightly bounded. This polygon does not consider that the road samples are just the average position taken from many samples, half of which would fall outside the lines created by the average sample positions. The area is better represented by dilating the polygons and using the expanded set of points to make a larger shape. Six points were taken at this radius around each point in the context area making a hexagon shape, with the intention that this would be sufficient to represent the expanded shape. The resulting points and shape can be seen in Figure 7. This technique applied over a larger area can be seen in Figure 12.

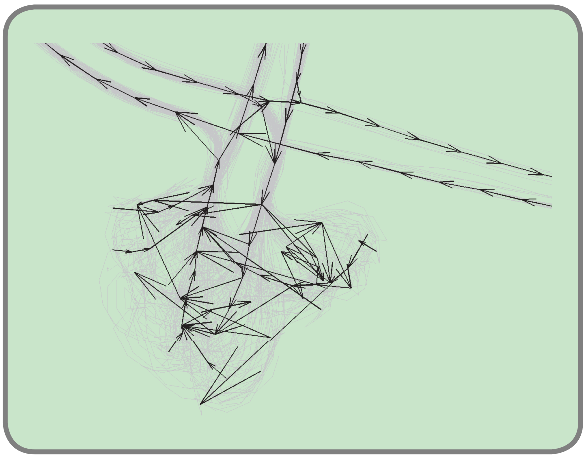


FIG 6 An area where vehicle motion is not well defined.

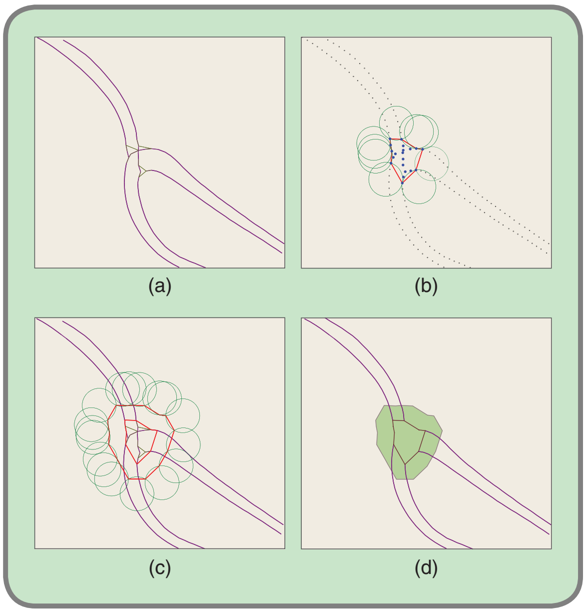


FIG 7 Using alpha shapes to bound an intersection.

III. Inferring Vehicle Dynamics by Considering Context

This section presents a probabilistic model for estimating the state of vehicles immersed in traffic and anticipate their future trajectories by reasoning about their interactions. The model is defined within a probabilistic filtering framework and represents interactions in terms of stochastic relationships between them.

The full state of an agent is divided into two distinct states: the dynamic state and context class. Low-level information such as position and orientation, velocity and acceleration comprises the *dynamic* state of the agent. Dynamics are low-level in the sense that they are directly observable through sensors such as global positioning system and inertial measurement units. The *context* class of an agent in its surroundings, on the other hand, is not entirely observable. Agent context is a symbolic representation of its state and encodes semantically meaningful information about the situation the agent is engaged in. For example, the location of the agent within a map or with respect to other agents are regarded as context variables.

At time t, the dynamic state and context class of an agent are, respectively, $\mathbf{x}_t$ and c_t. It is assumed that a predictive distribution exists over dynamic states, conditioned on the previous dynamic states and the context classes of all agents, as well as the current context class of the host agent. That is to say,

$$p\left(\mathbf{x}_t^{(a)} \middle| \bigcup_b \{\mathbf{x}_{t-1}^{(b)}, c_{t-1}^{(a)}\}, c_t^{(a)}\right), \tag{1}$$

where (a) indicates the ath agent and the union, denoted with (b) superscripts, is taken over all participating agents. Whereas dynamic states are typically continuous variables in Euclidean space, context classes are categorical variables that vary within a finite set $C = \{C_i\}$. The conditional probability that $c_t = C_i$ is modeled as a log-linear function of the features extracted from the previous states of all agents, that is

$$p\left(c_t^{(a)} = C_i \middle| \bigcup_b \{\mathbf{x}_{t-1}^{(b)}, c_{t-1}^{(b)}\}\right) \propto \exp\left(\mathbf{w}_i^T \mathrm{f}\left(\left\{\bigcup_b \mathbf{x}_{t-1}^{(b)}, c_{t-1}^{(b)}\right\}\right)\right). \tag{2}$$

Although it is not shown, for the sake of clarity, there is a normalization constant on the right-hand side of (2) that ensures that the probabilities sum to one—hence the proportionality sign.

The predictive distributions (1) and (2) make up the core of the model. While the former specifies how an agent moves according to its context, the latter connects the dynamic and context variables by way of linear combination of feature functions. Each feature term $\mathrm{w}_i^T \mathrm{f}$ maps the previous states of (potentially) all other agents into a real scalar that reflects the degree of compatibility between the dynamic states and the transition between context classes. Feature functions may have arbitrary shapes and hence endow the model with a great deal of flexibility. Both distributions will be described in more detail shortly.

In order to complete the model, it is necessary to define the way states relate to the data. Data are assumed generated directly from the dynamics. What this means is that, if the dynamic states of the agents were known, then data points would be statistically independent from their context. Additionally, each vehicle holds its own unique identifier and thus no data association ambiguities exist. The process that gives rise to the data is captured by a conditional probability density function $p(\mathbf{z}_t|\mathbf{x}_t)$, $\mathbf{z}_t$ being the data point received at time t for a given agent. The model is said to be generative since it explains the data generation mechanism; it relates agent dynamics to the data in a causal manner. For this reason, missing observations are readily dealt with. This is crucial when it comes to prediction—that is, propagating the state estimate forward in time—and during temporary sensor failures.

A. The Dynamic Layer

1) *Vehicle motion:* The dynamic state vector captures vehicle motion. The full dynamic state vector, $\mathbf{x}_t$, is composed of the horizontal and vertical position of the vehicle, x_t and y_t, its heading θ_t, velocity v_t and steering angle δ_t. The motion model is perturbed by additive white Gaussian noise,

$$\mathbf{x}_t \,|\, \mathbf{x}_{t-1}, c_t \sim \mathcal{N}(\mathrm{h}(\mathbf{x}_{t-1}, c_t), \mathbf{Q}(c_t)), \tag{3}$$

where the notation $\mathcal{N}(\mu, \Sigma)$ denotes a Gaussian distribution with μ its mean vector and Σ its covariance matrix.

The predictive mapping $\mathbf{h}$ and its gradient $\nabla\mathbf{h}$ are calculated as follows. The bicycle model, also known as the one-track model, details the way a vehicle of length Δl moves in terms of the differential equation,

$$\begin{cases} \dot{x}(t) = v(t)\cos\theta(t), \\ \dot{y}(t) = v(t)\sin\theta(t), \\ \dot{\theta}(t) = v(t)\tan\delta(t)/\Delta l, \\ \dot{v}(t) = a, \\ \dot{\delta}(t) = \omega, \end{cases} \tag{4}$$

where a and ω are, respectively, the linear acceleration and the turn rate. (These will be defined shortly as functions of the context). Denote by $\dot{\mathbf{x}}(t) = \mathbf{g}(\mathbf{x})$ the vector form of the differential equation in (4) above. Integrating this equation, together with its tangent map,

$$\begin{cases} \dot{\mathbf{x}}(t) = \mathbf{g}(\mathbf{x}(t)), \\ \dot{\mathbf{J}}(t) = \nabla\mathbf{g}(\mathbf{x}(t))^T\,\mathbf{J}(t), \end{cases}$$

over an interval of length Δt—with initial conditions $\mathbf{x}_0 = \mathbf{x}_{t-1}$, $\mathbf{J}_0 = \mathbf{I}$—yields the values $\mathbf{h}(\mathbf{x}_{t-1}, c_t)$ and $\nabla\mathbf{h}(\mathbf{x}_{t-1}, c_t)$, respectively.

Equation (1) sets forth the assumption that the dynamic layer is conditioned by the contextual layer. The process noise covariance matrix in equation (3) encodes this dependence and so does the linear acceleration and turn rate in equation (4),

$$a = a(c_t), \qquad \omega = \omega(c_t).$$

Consequently, each possible value of c_t has a corresponding $\mathbf{Q}(c_t)$, $a(c_t)$ and $\omega(c_t)$ that must be defined.

2) *Observation Model:* In this paper it is assumed that an absolute positioning system is the only source of data. The system generates a measurement $\mathbf{z}_t$ at each time step consisting of a latitude plus a longitude value. The observation model is defined as

$$\mathbf{z}_t|\mathbf{x}_t \sim \mathcal{S}(\mathbf{C}^T\mathbf{x}_t, \sigma^2\mathbf{I}, \nu), \tag{5}$$

where $\mathbf{C}$ is a rectangular matrix that selects the latitude and longitude entries from the dynamic state vector and concatenates them into a column vector. The symbol $\mathbf{I}$ denotes the identity matrix.

On the right-hand side of equation (5), the notation $\mathcal{S}(\mu, \sigma^2\mathbf{I}, \nu)$ stands for a Student-t distribution with mean μ, covariance $\sigma^2\mathbf{I}$ and ν degrees of freedom. The standard deviation σ is in units of meters and is inversely related to the accuracy of the positioning system. The number ν of degrees of freedom is a positive scalar and is introduced here to account for outliers in the data.

The Student-t is a sub-exponential distribution with much heavier tails than the Gaussian. Placing a Student distribution instead of the classic Gaussian provides additional robustness to outlying measurements by assigning them a non-negligible probability [12].

B. The Context Layer

The context of a vehicle is determined by its geometric and dynamic relationships with other vehicles in its vicinity. As an example, a driver who is stopped at a road intersection, waiting to join the main road, is likely to continue waiting for as long as there is incoming traffic. Only once traffic has cleared will he or she proceed to enter the junction. Feature functions capture the dynamic relationships influencing agent context. The set of all feature functions, linearly weighted according to their relative importance, compose the context layer.

1) *Feature functions and supervised classification:* Specifying the feature functions in (2) is a matter of design. They could, in principle, have virtually any mathematical form (on the condition that they are everywhere locally bounded). Any specific choice is problem-dependent and is usually motivated by the state representation. In the general case, many features must be defined so that context classes are linearly separable and thus classification is optimal.

Feature functions are often associated with supervised and semi-supervised classification. Within the supervised paradigm, input data are first labeled by an expert and then the classifier learns a mapping from inputs to classes, or posterior class probabilities, typically by fitting a set of linear weights. The criteria for adapting the weights are usually maximizing the conditional likelihood [13] or minimizing an empirical loss function [14], [15] of the data. If the objective function is convex (this is true for linear classifiers in combination with a negative log-likelihood criterion) then the optimal weight vector turns out to be unique.

In this case, data collected from vehicles are not labeled. Rather, the context classes are induced by the specific form of the feature functions and the context-conditional dynamic models. Just as for supervised classification, it is also possible to learn the set of weight vectors in (2) from the data, although under a different criterion: maximizing the expected log-likelihood of the data. Maximum-likelihood estimation of the weight vectors resembles the well-studied problem of fitting a multinomial logistic regression model. The logistic is a part of a more general class of models [16] that have been given a great deal of attention in the literature. Therefore, efficient parameter estimation methods abound. In particular, standard fitting tools include quasi-Newton methods such as the limited-memory algorithm of [17].

2) *Sufficient feature statistics:* Inference is often the most difficult part of parameter learning in hidden variable models. Inferring the posterior requires computing expected sufficient statistics involving the feature

functions. Namely, during filtering, the predicted context likelihood,

$$\tau_t^{(a)}(i) = E\left[p\left(c_t^{(a)} = C_i \middle| \bigcup_b \{\mathbf{x}_{t-1}^{(b)}, c_{t-1}^{(b)}\}\right)\right]$$
$$\propto E\left[\exp\left(\mathbf{w}_i^T \mathbf{f}\left(\left\{\bigcup_b \mathbf{x}_{t-1}^{(b)}, c_{t-1}^{(b)}\right\}\right)\right)\right] \qquad (6)$$

is evaluated intensely as it summarizes the influence exerted on agent (a) by all other participating agents. Numerical evaluation of these sufficient feature statistics is as follows. The expectations in (6) are taken with respect to the posterior over the previous dynamic states and context classes of all agents,

$$p\left(\bigcup_b \{\mathbf{x}_{t-1}^{(b)}, c_{t-1}^{(b)}\} \middle| \bigcup_b \bigcup_{k=1}^{t-1} \mathbf{z}_k^{(b)}\right) = \prod_b p\left(\mathbf{x}_{t-1}^{(b)}, c_{t-1}^{(b)} \middle| \bigcup_c \bigcup_{k=1}^{t-1} \mathbf{z}_k^{(c)}\right), \qquad (7)$$

a distribution that factors across agents. Due to the above factorization, sampling from (7) equates to drawing a sample from each marginal in turn. This allows (6) to be evaluated efficiently via Monte Carlo quadrature [18].

The factorization (7) stems from the coupling between agents being defined across succeeding time slices. It is a direct consequence of (1) and (2) and is readily derived by applying the fundamental rules of Bayesian probability. If, on the contrary, the predictive distribution (2) were conditioned by variables at time t instead of $t-1$, then (7) would no longer hold.

The factored structure of (7) is also a consequence of the fact that inference consists of filtering only. In other words, the only quantity of interest is the filtered distribution,

$$\alpha_t^{(a)}(c_t^{(a)}, \mathbf{x}_t^{(a)}) \triangleq p\left(c_t^{(a)}, \mathbf{x}_t^{(a)} \middle| \bigcup_b \bigcup_{k=1}^{t} \mathbf{z}_k^{(b)}\right), \qquad (8)$$

which is updated as new data arrive. Performing state smoothing would destroy the factorization since agent correlations propagate backwards in time, meaning that the past state of an agent comes to be dependent on the future states of the others.

3) *Learning the weight vectors:* The weight vectors are learnt according to the maximum likelihood criterion by applying the Expectation Maximization (EM) algorithm of [19] (see also [20]). The EM algorithm is established in problems such as estimating the parameters of a statistical model involving hidden variables. It proceeds by maximizing the marginal log-likelihood of the data in a coordinate-wise manner by alternating between the Expectation (E) step and the Maximization (M) step. The E step infers the posterior distribution over hidden variables and collects the relevant expected sufficient statistics. The M step updates the weight vectors in order to maximize, or at least increase, the expected complete-data log-likelihood.

For a set of m data, the expected complete-data log-likelihood is given by

$$\sum_a \sum_{t=2}^{m} \mathrm{E}\left[\ln p\left(c_t^{(a)} \middle| \bigcup_b \{\mathbf{x}_{t-1}^{(b)}, c_{t-1}^{(b)}\}\right)\right], \qquad (9)$$

expectations taken with respect to the posterior inferred from the data during the E step. Because the maximum of (9) has no closed-form solution, the M step adjusts the weights according to the gradient. Specifically, the weights and the direction of ascent are updated by a limited-memory version of the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm [17]. BFGS requires the gradient of the complete-data log-likelihood with respect to the jth weight vector $\mathbf{w}_j$, which is

$$\sum_a \sum_{t=2}^{m} \sum_{i=1}^{|C|} \mathrm{E}\left[\rho_t^{(a)}(i,j)\, \mathbf{f}\left(\bigcup_b \{\mathbf{x}_{t-1}^{(b)}, c_{t-1}^{(b)}\}\right)\right],$$

where

$$\rho_t^{(a)}(i,j) = p\left(c_t^{(a)} = C_j \middle| \bigcup_b \{\mathbf{x}_{t-1}^{(b)}, c_{t-1}^{(b)}\}\right) - \delta(i,j).$$

It can be shown that (9) is convex with respect to the weights.

C. Statistical Inference

The purpose of inference, for the model introduced in this manuscript, is to track the distribution over dynamic and contextual states of all the agents as they develop through time. There are a number of possible alternatives for the inference algorithm, and a careful selection is essential as it affects the performance and scalability of the model.

1) *The inference engine*: Exact inference in the model is, in general, neither analytically nor computationally tractable. The combination of continuous (dynamic) and discrete (contextual) variables causes the computational complexity to increase exponentially with time. This is infeasible in practice if computer resources are limited. Further, the coupling between the dynamical and contextual layers originating from the feature functions requires expectations of arbitrary functions to be evaluated with respect to continuous distributions. These expectations require integration and there may be no closed-form analytic solution in the general case.

Approximate inference is tackled on a per-agent basis with an extended version of the forward algorithm of [21], [22] for Augmented Switching Linear-Dynamical Systems (ASLDS). Extensions include accounting for the non-linear nature of the predictive function and the non-Gaussian observation model, as well as confining the computational complexity of the posterior distribution. Prediction under the non-linear function is approximated by linearizing it as in the extended Kalman filter. Incorporating the Student-t

distribution in (5) is achieved by way of a structured variational approximation [23], [24], along the lines of [12], which results in a direct modification of the standard Kalman update step. Last of all, iteratively reducing the posterior—approximated as a mixture distribution—keeps the computational complexity within bounds. A Gaussian mixture can approximate any smooth distribution to an arbitrary accuracy [25]. In addition, principled approaches based on the integrated squared distance [26] or the Kullback-Leibler divergence [27] exist for mixture reduction.

Although the causes of accidents in haul trucks for surface mining are numerous, the most important one is often the operator's lack of situational awareness.

IV. Experimental Results

Safety in surface mining is an important and growing area of application for advanced driving assistance systems. Year after year, accidents take place involving impending interactions between haul trucks and other resources such as light vehicles, loaders and graders. In addition to many of them resulting in serious injury and even fatality, they lead to significant losses in equipment and incur large costs due to repair and downtime. Although the causes are many, the most important one is often the operator's lack of situational awareness [28], [29]. Surface mines are complex and hostile environments. Heavy machinery and light vehicles constantly operate and interact at close range. Visibility is extremely limited by the size of the machinery, which creates blind areas in his or her field of view, and may be severely impaired by harsh environmental conditions such as fog, dust, rain and snow.

The approach presented here lends itself to the development of advanced driving assistance systems. The main purpose of these systems is to generate proper feedback for the vehicle operator, thereby enabling them to deal with traffic safely and comfortably. Generating just enough feedback is critical since, most of the time, interactions do not represent a threat that warrants an alarm. An overly conservative system that raises too many false alarms will most likely end up being counter productive as it will overburden the driver with useless information.

This section shows results for the mapping and the dynamic state and context class inference algorithms. Position data from a fleet of surface mining vehicles were collected at an opencast mine in Western Australia. Figure 8 illustrates the mining trucks that collected the data. The antennas required by the system are depicted at the top-front of the trucks. The data span one week of operation of the mine. During this time several resources, including haul trucks, shovels and light vehicles, participated in the excavation process and interacted with each other. Data were recorded for seven haul trucks that took part in the operation. Position was logged with standard non-differential GPS units working in autonomous mode at an average rate of $\Delta t = 1\ s$. Data were collected from the vehicles automatically, and stored in a database as illustrated in Figure 9.

FIG 8 System installed on mining vehicles.

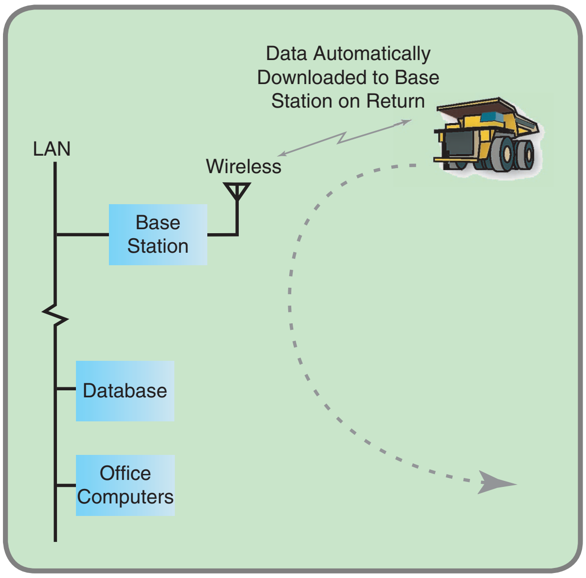


FIG 9 Data is automatically collected from vehicles during operation.

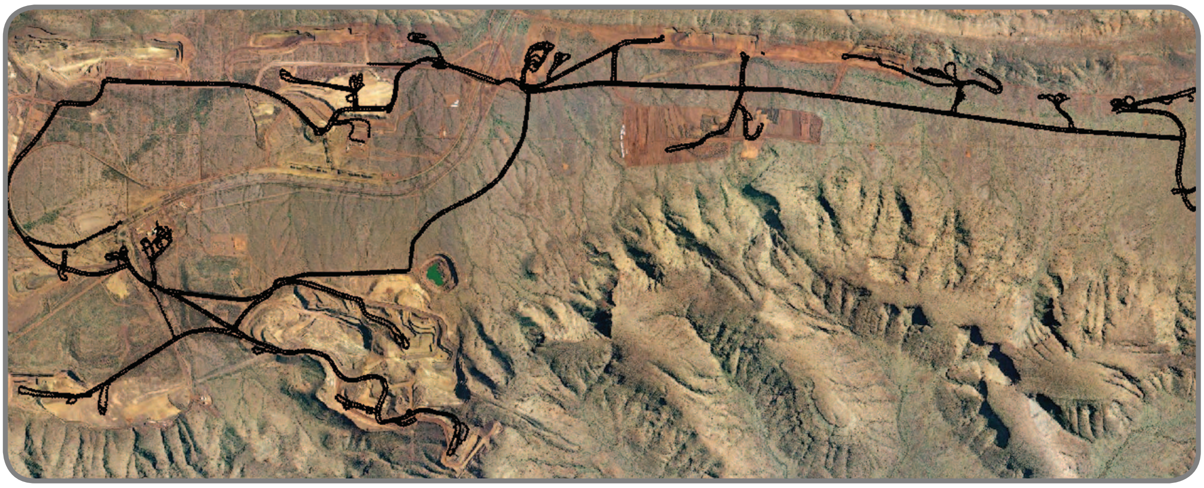


FIG 10 Final road map superimposed on an aerial photograph of the mine. The map is drawn in black. Although details may not be appreciable at this scale, it still serves to contemplate the magnitude of the mining activity. The photograph spans an area of 10.5 by 3.5 kilometers.

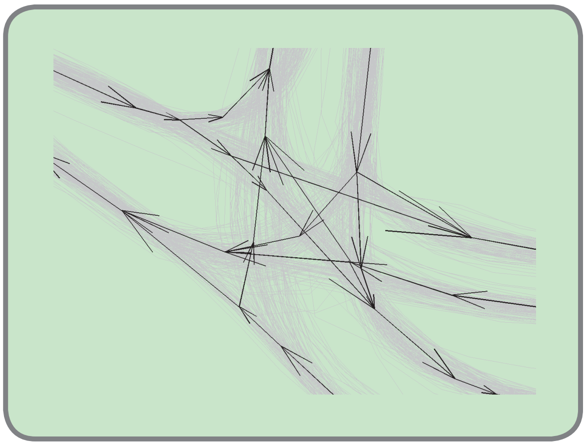


FIG 11 Linking sample points in a complex intersection.

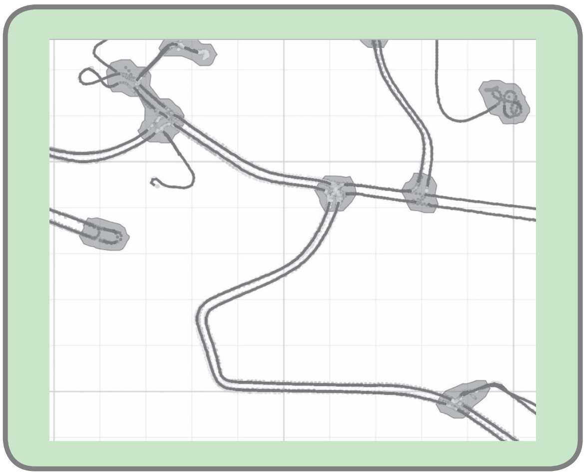


FIG 12 A map showing many bounded areas of interest. Coarse grid overlay represents 500 meter squares.

A. Mapping and the Location Context

The road network and set of area polygons was estimated from the position data collected during vehicle operation. The sampling and linking process successfully extracted the structure of the road network, even in intersections with a large degree of clutter. Figure 10 illustrates the final road map superimposed on an aerial image of the mine. Some typical intersections are seen in Figures 4 and 11. Notice particularly in 11 that there is a large lateral variation in the vehicle trajectories for vehicles travelling from south to north. Despite this variation, however, the algorithm is still able to extract the correct road network structure.

An example of the algorithm output for bounding the areas of importance is shown in Figure 12. Position information is plotted with the non-intersection roads and polygons. The polygons cover the position data that are not assigned to the non-intersection road samples. The majority of the polygon clearly delimit an area that is likely to be interesting, such as an intersection, a curve or a loading area. It is important to note that two polygons in close proximity will be merged together, as on the top-left of Figure 12. A smaller dilation distance of the bounding polygon (from Section II-D) would result in two separate polygons. Whether the area is best represented by one or more separate polygons is subjective, and represents consideration for future work.

B. Dynamic States and Context

The vehicle dynamic model presented in Equation (1) was estimated from the data. Assuming $a = \omega = 0$, a nominal parameter set $\mathbf{Q}, \sigma, \nu$ was fit to a few hundred points from the available data, with the unsupervised learning method of [12]. After learning, the $\mathbf{Q}$ matrix was scaled and a set of three contexts, labelled as "braking," "constant velocity" and "accelerating," were defined as

$$\begin{cases} a(c_1) = -1, & \text{(braking)} \\ a(c_2) = 0, & \text{(constant velocity)} \\ a(c_3) = +1, & \text{(accelerating)} \end{cases}$$

and $\omega(c_t) = 0$. All units are given in m/s^2.

Three feature functions were defined. The first is intended to impose temporal consistency across consecutive time steps and was defined as 1 if $c_{t-1} = c_t$ and zero otherwise. The second relates the dynamical and contextual layers; it checks whether the change in velocity is below -1 or above $+1$ m/s, and emits a one or a zero according to the current context (e.g., for "braking," the function evaluates to 1 if the velocity decreases by more than 1 m/s^2). Finally, the third feature function couples the two agents together; it is 1 for the agent that is travelling slowest of all. Although in principle the feature weights could be learnt from the data, they were set to one for the purpose of illustration.

The algorithm presented in Section III was implemented in MatLab® and tested on an Intel® Core™ Duo CPU with a 2.33 GHz processor and 2 Gb of RAM. For a mixture size of five components, each filtering step took an average of 0.6 seconds (including iteration until convergence).

The Gaussian mixture components are plotted in figure 13. Covariance ellipses for each of the components are drawn with transparency proportional to the component weight. The marginal distributions over contexts can be seen in figure 14. Solid lines denote the context probabilities for agent 17, whereas dashed lines correspond to agent 35. Both figures 13 and 14 are aligned so that they span approximately the same time interval.

It is interesting to observe how the posterior distribution evolves through time. As agent 35 approaches the main haul road and brakes to give way to agent 17, the data becomes noisier, since GPS measurements are less accurate at lower speeds. Therefore, the uncertainty ellipses tend to widen. Notice, however, that the model remains confident about the contextual states of the vehicles and correctly infers that agent 35 yields to agent 17. Although this is a fairly simple example, it serves to illustrate the potential of the approach.

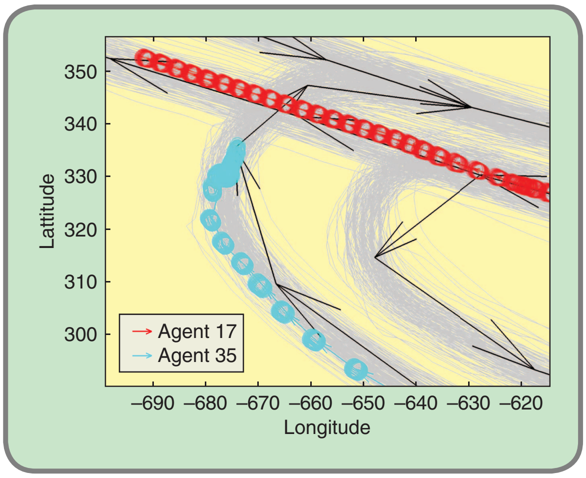


FIG 13 Filtering results superimposed on the interaction data.

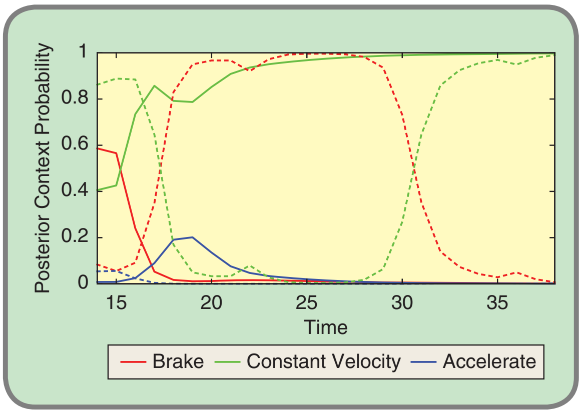


FIG 14 Posterior probability over dynamic contexts for each agent.

V. Conclusions

This paper introduced the Intelligent Systems for Risk Assessment (ISRA) project, which aims to reduce accidents in vehicle operations by using context based risk analysis. The importance of using context information was highlighted and different types of context information were introduced.

A process for learning the location context information (roads and areas) was presented. The process of sampling from available data, linking and determining important areas was demonstrated using real data collected from a fleet of mining vehicles during normal operation. The process outlined in this paper was successful in determining the structure of the road network including intersections, and successfully bound non-structured areas into "important areas."

A method for inferring the high level vehicle behavior using context information was also presented. This process was able to infer using vehicle state information and context information that one vehicle was yielding to another. Feature functions were used to encapsulate expert knowledge, and future work will examine introducing and learning new functions.

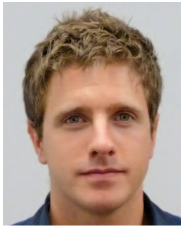
The work presented in this paper represents the first stage of the project that aims to measure risk of accidents based on the context. By extending the location and behavior context information to include more complex scenarios such as multiple lanes and complex intersections, it is expected that the ISRA project will perform context based risk detection in the ITS domain. The risk analysis metrics can be used to prevent accidents by identifying high risk scenarios and providing graphical and/or audible warnings to the operator, or potentially take control of the vehicle in the case an unacceptably level of risk is detected.

VI. Acknowledgments

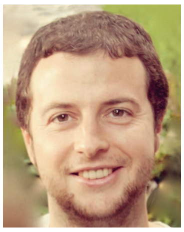
This work was partly supported by the Australian Centre for Field Robotics and the Rio Tinto Centre for Mine Automation, University of Sydney.

About the Authors

Stewart Worrall completed a Ph.D. at the University of Sydney, Australia in 2009. He is currently a research associate at the Australian Centre for Field Robotics at the University of Sydney. His research is focused on the study and application of technology for vehicle automation and improving vehicle safety.



Gabriel Agamennoni received his bachelor degree in Electronic Engineering at the Universidad Nacional del Sur, Argentina. He began his Ph.D. candidature at the Australian Centre for Field Robotics at the University of Sydney in 2008. His research interests are statistical modeling and machine learning.



Juan Nieto received his Ph.D. in robotics at the University of Sydney in 2005. Until 2007 he was a Research Associate at the Australian Centre for Field Robotics. From 2007 he has been a Senior Research Fellow at the Rio Tinto Centre for Mine Automation, University of Sydney. His main research interest is in navigation systems and perception.

Eduardo Nebot, B.Sc. EE, UNS, Argentina and M.S. and Ph.D. CSU, USA. He is a Professor at the University of Sydney and the Director of the Australian Centre for Field Robotics. His main research areas are in Field Robotics Automation. The major impact of his fundamental research is in autonomous system, navigation and mining safety.

References

[1] T. Michalke and R. Kastner, "The attentive co-pilot: Towards a proactive biologically-inspired advanced driver assistance system," *IEEE Intell. Transport. Syst. Mag.*, pp. 1–18, 2011.

[2] D. Hendricks, U. S. N. H. T. S. Administration, and I. Veridian Engineering, *The Relative Frequency of Unsafe Driving Acts in Serious Traffic Crashes*. Washington, DC: National Highway Traffic Safety Administration, 2001.

[3] U. Ozguner, C. Stiller, and K. Redmill, "Systems for safety and autonomous behavior in cars: The DARPA grand challenge experience," *Proc. IEEE*, vol. 95, no. 2, pp. 397–412, Feb. 2007.

[4] C. Urmson and W. Whittaker, "Self-driving cars and the urban challenge," *IEEE Intell. Syst.*, vol. 23, no. 2, pp. 66–68, Mar.–Apr. 2008.

[5] I. Dagli and D. Reichardt, "Motivation-based approach to behavior prediction," in *Proc. IEEE Intelligent Vehicle Symp.*, 2002, pp. 227–233.

[6] D. Meyer-Delius, C. Plagemann, G. von Wichert, W. Feiten, G. Lawitzky, and W. Burgard, "A probabilistic relational model for characterizing situations in dynamic multi-agent systems," in *Proc. 31st Annu. Conf. Data Analysis, Machine Learning and Applications*, 2007, pp. 269–276.

[7] T. Gindele, S. Brechtel, and R. Dillmann, "A probabilistic model for estimating driver behaviors and vehicle trajectories in traffic environments," in *Proc. IEEE Int. Conf. Intelligent Transportation Systems*, 2010, pp. 1625–1631.

[8] C. Pantilie, S. Bota, I. Haller, and S. Nedevschi, "Real-time obstacle detection using dense stereo vision and dense optical flow," in *Proc. IEEE Int. Conf. Intelligent Computer Communication and Processing (ICCP)*, 2010, pp. 191–196.

[9] G. Agamennoni, J. Nieto, and E. Nebot, "Robust inference of principal road paths for intelligent transportation systems," *IEEE Trans. Intell. Transport. Syst.*, vol. 12, no. 1, pp. 298–308, Mar. 2011.

[10] S. Worrall, D. Orchansky, F. Masson, J. Nieto, and E. Nebot. (2010). Determining high safety risk scenarios by applying context information, *J. Phys. Agents* [Online]. *4(2)*. Available: http://www.jopha.net/index.php/jopha/article/view/70

[11] H. Edelsbrunner and E. Mücke, "Three-dimensional alpha shapes," in *Proc. workshop on Volume visualization*, ACM, 1992, p. 82.

[12] G. Agamennoni, J. Nieto, and E. Nebot, "An outlier-robust Kalman filter," in *Proc. IEEE Int. Conf. Robotics and Automation*, 2011.

[13] M. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, 2001.

[14] K. Crammer and Y. Singer, "On the algorithmic implementation of multi-class kernel-based vector machines," *J. Mach. Learn. Res.*, vol. 2, pp. 265–292, 2001.

[15] B. Taskar, C. Guestrin, and D. Koller, "Max-margin Markov networks," in *Advances in Neural Information Processing Systems*. Cambridge, MA: MIT Press, 2003.

[16] J. Nelder and R. Wedderburn, "Generalized linear models," *J. Royal Stat. Soc. A (General)*, vol. 135, no. 3, pp. 370–384, 1972.

[17] R. Byrd, P. Lu, J. Nocedal, and C. Zhu, "A limited-memory algorithm for bound-constrained optimization," *SIAM J. Scientific Comput.*, vol. 16, no. 5, pp. 1190–1208, Sept. 1995.

[18] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, Oct. 1992.

[19] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Royal Stat. Soc. B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977.

[20] R. Neal and G. Hinton, "A view of the EM algorithm that justifies incremental, sparse, and other variants," in *Proc. NATO Advanced Study Institute on Learning in Graphical Models*, 1998.

[21] D. Barber, "Expectation correction for smoothed inference in switching linear dynamical systems," *J. Mach. Learn. Res.*, vol. 7, pp. 2515–2540, Dec. 2006.

[22] B. Mesot and D. Barber, "A simple alternative derivation of the expectation correction algorithm," *IEEE Signal Processing Lett.*, vol. 16, no. 2, pp. 121–124, Feb. 2009.

[23] T. Jaakola, "Tutorial on variational approximation methods," in *Advanced Mean Field Methods: Theory and Practice*, M. Opper and D. Saad, Eds. Cambridge, MA: MIT Press, 2000.

[24] M. Beal, "Variational algorithms for approximate Bayesian inference," Ph.D. dissertation, Gatsby Computational Neuroscience Unit, Univ. College London, 2003.

[25] V. Maz'ya and G. Schmidt, "On approximate approximations using Gaussian kernels," *IMA J. Numerical Anal.*, vol. 16, no. 1, pp. 13–29, 1996.

[26] J. Williams, "Cost-function-based Gaussian mixture reduction for target tracking," in *Proc. 6th Int. Conf. Information Fusion*, vol. 2, 2003, pp. 1047–1054.

[27] A. Runnalls, "Kullback-Leibler approach to Gaussian mixture reduction," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 43, no. 3, pp. 989–999, July 2007.

[28] S. Worrall and E. Nebot, "Using non-parametric filters and sparse observations to localise a fleet of mining vehicles," in *Proc. IEEE Int. Conf. Robot. Automat.*, Apr. 2007, pp. 509–516.

[29] S. Worrall and E. Nebot, "A probabilistic method for detecting impending vehicle interactions," in *Proc. IEEE Int. Conf. Robot. Automat.*, May 2008, pp. 1787–1791.

ITS