Graphical Models for Driver Behavior Recognition in a SmartCar

Nuria Oliver *Alex P. Pentland*

Media Laboratory
Massachusetts Institute of Technology (MIT)
20 Ames St E15-385 Cambridge MA 02139
Ph: 617 253 0608 - Fax: 617 253 8874

nuria@media.mit.edu *sandy@media.mit.edu*

Abstract

In this paper we describe our SmartCar testbed: a real-time data acquisition system and a machine learning framework for modeling and recognizing driver maneuvers at a tactical level, with special emphasis on how the context affects the driver's performance. The perceptual input is multi-modal: four video signals capture the contextual traffic, the driver's head and the driver's viewpoint; and a real-time data acquisition system records the car's brake, gear, steering wheel angle, speed and acceleration throttle signals. Over 70 drivers have driven the SmartCar for 1.25 hours in the greater Boston area. Graphical models, HMMs and potentially extensions (CHMMs), have been trained using the experimental driving data to create models of seven different driver maneuvers: passing, changing lanes right and left, turning right and left, starting and stopping. We show that, on average, the predictive power of our models is of 1 second **before** *the maneuver starts taking place. Therefore, these models would be essential to facilitate operating mode transitions between driver and driver assistance systems, to prevent potential dangerous situations and to create more realistic automated cars in car simulators.*

1 Introduction and Background

In this paper we describe our SmartCar testbed: a real-time data acquisition system and a machine learning framework for modeling and recognizing driver maneuvers at a tactical level, with special emphasis on how does the context affect the driver's performance. The perceptual input is multi-modal: four video signals capture the contextual traffic, the driver's head and the driver's viewpoint; and a real-time data acquisition system records the car's brake, gear, steering wheel angle, speed and acceleration throttle signals. Over 70 drivers have driven the SmartCar for 1.25 hours in the greater Boston area. Graphical models, HMMs and extensions, are trained using the experimental driving data to create models of seven different driver maneuvers: passing, changing lanes right and left, turning right and left, starting, and stopping. The accuracy of these models in recognizing the driving maneuvers is reported. More importantly, our models accurately recognize the next driving maneuver on average 1 **second before** the maneuver starts taking place.

Human driver modeling is an interdisciplinary endeavor involving a number of fields including robotics, psychology, control theory and statistics. Driving in a real-life traffic situation is a very difficult task because good decisions need to be made given only incomplete information in real time.

One critical issue in machine-human interface systems are the transitions between manual and automated operation. They should be as seamless and smooth as possible. It is critical not to interfere with the driver's intended maneuver, specially in emergency situations, and to avoid discontinuities in the system, inducing feelings of incongruity while driving. Therefore, developing systems for predicting the driver's next maneuver or inferring driver's intentions is imperative to facilitate smooth and appropriate control mode transitions.

Building effective driver behavior recognition methods requires a thorough understanding of driver behavior and the construction of a model capable of both generating and explaining the drivers' behavioral characteristics. In this paper we focus on driving maneuvers at a tactical level. Namely, we have built models of seven maneuvers from data in real driving situations: passing, changing lanes right and left, turning right and left, starting, and stopping.

Most of the driver behavior models previously proposed in the literature have not been human model-based. The most closely related work to ours is that of Pentland and Liu [5], and that of Kuge at al [2]. In [5] Pentland and Liu develop a computational state-based model of driver behavior using Hidden Markov Models (HMMs). Their system is able to predict when the driver is about to brake or turn. Kuge et al. present a similar HMM method that characterizes and detects lane changing maneuvers.

None of these previous systems, however, incorporates contextual information when modeling driver behavior. Nonetheless, knowledge of the context is necessary to properly make decisions in complex dynamic environments such as driving. Psychologists attribute this competence to a task-specific understanding of the situation, termed *situation awareness*. In this paper we develop machine models of driver behavior that incorporate elements of situational awareness for tactical driving. Moreover perceptual information is critical to driver behavior modeling. For example, recent research shows that drivers' eye fixation patterns are strongly correlated with their current mental state. The work presented in this paper combines perceptual modules with behavior modules in a closed-loop.

There is today strong research efforts invested in developing partially or fully automated driver assistance systems. For example, headway distance control or lane keeping control systems, which make use of Intelligent Transportation System (ITS) technologies. To achieve such assistive systems, it is important to adopt approaches aimed at improving the performance of the driver-vehicle-context cooperative system by regarding driving as an interaction between the driver, the vehicle, the road information and surrounding traffic.

Finally, it has also been argued that laboratory research of SA should be conducted under conditions that afford as much realistic behavior as possible. Due to the simplicity of most car simulators, specially the lack of realism of the computer generated automated cars, the experiments carried out in this paper took place in a real car while driving in the greater Boston area.

In this paper we extend Pentland and Liu's framework [5] in several ways: (1) we model a larger number of maneuvers at a tactical level –namely seven–; (2) we show that contextual information is critical for the accurate recognition of some maneuvers; (3) we use real data collected in an instrumented car, as opposed to using a car simulator.

The rest of the paper is structured as follows: section 2 presents an overview of the system; the perceptual input of the SmartCar testbed is described in section 3; the statistical models used for behavior modeling and recognition are described in section 4. Section 5 contains the description of our experiments and reports the recognition results in real driving situations. Finally, section 6 summarizes the main conclusions and outlines our future lines of research.

2 System's Architecture

Figure 2 shows the system's architecture. In the proposed architecture, there is a *bottom-up* stream of information gathered with the various sensors, and a *top-down* information flow through the predictions provided by the behavior models. Our mathematical framework follows a Bayesian approach such that both the observations (bottom-up) and complex behavioral priors (top-down) can be combined in a sound way, to provide expectations that could eventually be fed back to the perceptual system.

3 Perceptual Input

There are at least three different elements beyond the driver that are relevant when driving at a tactical level:

1. SmartCar physical self-state: information sensed from the speedometer, acceleration throttle, steering wheel angle sensor, brake pedal, gear and GPS unit.
2. Road state: including road geometry and exit information.
3. Traffic state: relative speeds and positions of the surrounding cars.

The sensors installed in the SmartCar provide information about (1) the car's internal state (brake, acceleration throttle, steering wheel angle, gear and speed), (2) surrounding traffic and lanes (via two Sony EVI-D30 cameras with wide field of view, mounted on the front dashboard and on a tripod in the trunk, to record frontal and rear traffic respectively), and (3) driver's face and gaze position, and driver's viewpoint (by use of an ELMO CCD camera mounted on the steering wheel, recording the driver's face, and another ELMO CCD camera mounted on a pair of glasses to record the driver's viewpoint). All the video signals are combined in a quadsplitter whose output is recorded using a Sony GV-A500 Hi8 Video Walkman VCR.

We have developed the hardware and software for acquiring in real-time car state data. All the signals but the steering wheel are available directly from the

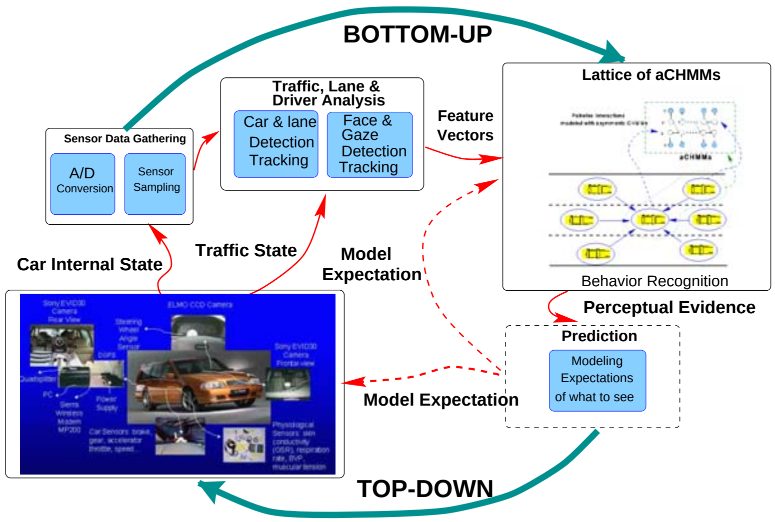

Figure 1: SmartCar architecture

car electronics system. We have designed a steering wheel angle sensor and mounted it on the car's steering wheel. The hardware obtains its inputs from sources of three different nature as shown in table 1. All the car signals are connected to a PCMCIA Data Acquisition Card by National Instruments. The analog signals are digitized and sampled at 150 scans/s. The digital signals are also sampled at 150 scans/s. All the signals can be directly connected to one of the DAQ boards, except for the speed, because it consists of a 12 pulse-per-revolution signal. Therefore we designed a frequency-to-voltage converter to first convert it to analog. The software for data acquisition, processing

Signal	Nature	Description
Speed	Analog	12 pulse/wheel rev. square wave
Acc	Analog	Linear 0-12 V
Brake	Digital	Boolean (0=off, 1=on)
Gear	Digital	2-bit
Steering angle	Analog	Up to 3 revolutions (variable voltage) (in a potentiometer)

Table 1: Sensor signals in the SmartCar.

and playback has been developed in LabVIEW. We have also developed a graphical user interface for calibrating the car signals, triggering the acquisition, and annotating the driving maneuvers as they take place.

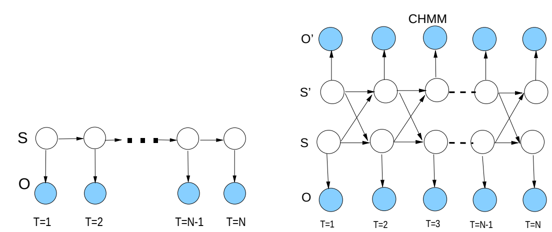

Figure 2: Graphical representation of a HMM and a CHMM rolled-out in time

4 Driver Behavior Models

Statistical directed acyclic graphs (DAGs) or probabilistic inference networks (PINs) can provide a computationally efficient solution to the problem of time series analysis and modeling. HMMs and some of their extensions, in particular CHMMs [1, 3], can be viewed as a particular and simple case of temporal PIN or DAG. Graphically Markov Models are often depicted 'rolled-out in time', such as in figure 2.

PINs present important advantages that are relevant to our problem: they can handle incomplete data as well as uncertainty; they are trainable and easier to avoid overfitting; they encode causality in a natural way; there are algorithms for both doing prediction and probabilistic inference; they offer a framework for combining prior knowledge and data; and finally they are modular and parallelizable.

Traditional HMMs offer a probabilistic framework for modeling processes that have structure in time. They offer clear Bayesian semantics, efficient algorithms for state and parameter estimation, and they automatically perform dynamic time warping. An HMM is essentially a quantization of a system's configuration space into a small number of discrete states, together with probabilities for transitions between states. A single finite discrete variable indexes the current state of the system. Any information about the history of the process needed for future inferences must be reflected in the current value of this state variable.

However many interesting real-life problems are composed of multiple interacting processes, and thus merit a compositional representation of two or more variables. This is typically the case for systems that have structure both in time and space. With a single state variable, Markov models are ill-suited to these problems. In order to model these interactions a more complex architecture is needed. We have developed a new architecture called Coupled Hidden Markov Models (CHMMs) for modeling interacting processes [1, 4], such as cars.

The posterior state sequence probability $P(S|O)$ for HMMs is given by

$$P(S|O) = P_{s_1} p_{s_1}(o_1) \prod_{t=2}^{T} p_{s_t}(o_t) P_{s_t|s_{t-1}} \quad (1)$$

where $S = \{a_1, \ldots, a_N\}$ is the set of discrete states, $s_t \in S$ corresponds to the state at time t. $P_{i|j} \doteq P_{s_t=a_i|s_{t-1}=a_j} = P_{s_t|s_{t-1}}$ is the state-to-state transition probability (i.e. probability of being in state a_i at time t given that the system was in state a_j at time $t-1$). $P_i \doteq P_{s_1=a_i} = P_{s_1}$ are the prior probabilities for the initial state. Finally $p_i(o_t) \doteq p_{s_t=a_i}(o_t) = p_{s_t}(o_t)$ are the output probabilities for each state[1]. The well-known Baum-Welch algorithm efficiently estimates – using dynamic programming– the state posterior probability in an HMM (inference problem). Just as with the inference problem, the Viterbi algorithm provides an efficient, locally recursive method for solving the MAP identification problem with complexity TN^2.

5 Experiments

Apparatus

An instrumented automatic Volvo V70XC was used to gather driver behavior data. The car sensors have been described in section 3.

The procedure The driving task took place in the greater Boston area. Over 70 drivers drove both in the city and in different highway sections, for about 1.5 hours. The drivers were asked to sign an consent form before starting the experiment. They were rewarded $20 for participating. After the driving task was completed, the drivers were asked to fill in a questionnaire with basic questions about their driving experience, skills and the experiment.

A driving instructor was with the driver throughout the experiment. The instructor gave directions to the driver about where to go and labeled the driving maneuvers as they took place using the laptop computer and the LabVIEW GUI. Because our focus is on predicting what is the most likely maneuver to take place next, the driver was requested to verbally report his/her **next intended action before** carrying it out. The four video signals were recorded for the entire route. The car signals, however, were only recorded whenever a maneuver was happening. A time window of 2 seconds was used, such that we started recording the car signals 2 s **before** the driver reported his intentionality to perform a maneuver. Both the video and car data was time-stamped (the VCR and the laptop clocks were synchronized before every session). We collected data of: passing, turning right/left, changing lanes right/left, starting, stopping and merging.

Figures 3 and 4 show typical car and context signals in one example of a 'passing' maneuver collected in the experiments. Note how, in the case of passing, the car signals contain little information about the maneuver type, whereas the gaze and lane are much more relevant features.

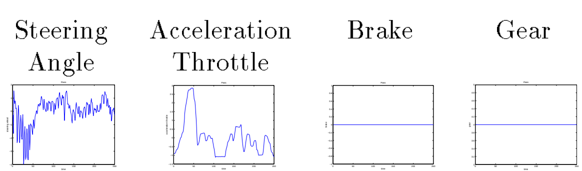

Figure 3: Typical car signals for a passing maneuver

Data post-processing and Driver Models The contextual information was acquired via the video signals. We have developed a video processing graphical environment that let's the user record, playback and annotate the video signals coming from the front, rear and face driver cameras. Contextual information (the driver's gaze, the relative position of the road lanes or the relative position, velocities and direction of the surrounding traffic) was manually annotated for each frame and maneuver.

Due to the different sampling rate on the car and video

[1] The output probability is the probability of observing o_t given state a_i at time t

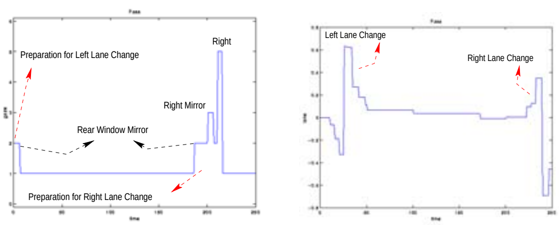

Figure 4: Typical contextual signals for a passing maneuver: gaze (left) and lane feature (right)

signals, we subsampled the car data to match the video frame rate. The final sampling rate was of approximately 30 samples/s. All the continuous signals were low-pass filtered using Butterworth filters.

Using the car, driver's gaze and road lane data, we built HMMs for each of the maneuvers to be recognized. The number of states in the HMMs was determined using 10-fold cross-validation. We evaluated the performance on recognition (accuracy) of the best HMMs trained with different feature vectors: (1) Car data: brake, steering wheel angle, gear, and acceleration throttle; (2) car data and lane position information; (3) car data and driver gaze information; (4) car data, lane and driver information.

The gaze was a discrete signal with 6 possible values: (1) front road, (2) rear window mirror, (3) right mirror, (4) left mirror, (5) right and (6) left.

In the case of the lanes, a single value (lane feature) was computed from the (x, y) image coordinates of the road front/back right/left lanes:

$$lane_{feat} = \frac{lane_{fr} + lane_{br} - (lane_{fl} + lane_{bl})}{4.0} \quad (2)$$

where $lane_i = \mathrm{atan2}(|y_2 - y_1|, |x_2 - x_1|)$, with $i \in$ {front left (fl), front right (fr), back left (bl) and back right (br)}. The best models (best number of states and feature vector) were selected using 10-fold cross-validation. The training data set was about 80% of the total amount of data. The testing data set consisted of the rest of the data that had not been used for training.

The number of examples collected in the driving experiments is summarized in table 2. The average length of the longest maneuver (passing) was 15 s and the shortest (lane changes and starting) was 4.5 s.

The results on recognizing the previous driving maneuvers are depicted in table 3. Some interesting conclusions to be drawn from the experimental results are:

	Car data	Traffic data
Pass	710	40
turn right	257	37
turn left	260	31
change lane right	663	81
change lane left	711	87
start	401	30
stop	404	26

Table 2: Number of driving examples

Accuracy (%)			
	Car	Car & Lane	Car & Gaze
pass	100.0	100.0	100.0
turn right	71.4	85.7	85.7
turn left	0.0	33.3	66.7
change lane right	0.0	12.5	6.3
change lane left	29.4	17.6	23.5
start	100.0	66.7	83.3
stop	100.0	100.0	100.0

Table 3: Accuracy for HMMs car only, car and lane and car and gaze data

1. There is a plateau of accuracy that can be reached using car information only. Certain maneuvers – such as passing and changing lanes left– cannot be accurately distinguished using car information only.

2. The context is crucial for recognizing maneuvers such as turnings and lane changes.

3. The driver's gaze seems to be strongly correlated with the driver's mental state in real life driving. It is, thus, a relevant feature for driver maneuver prediction, specially in lane changes, passings and turnings.

4. **Predictive Power**: The models are able to recognize the maneuver on average 1 **second before** any significant (20% deviation) change in the car or contextual signals take place. Table 4 contains the average prediction power for each of the maneuvers, and figure 5 illustrates through an example what this *predictive power* means. It depicts, frame by frame, the lane feature and the -log(likelihood) of the different models for a passing maneuver. There is no significant change in the lane position until frame 26. However, the models are able to recognize the passing from

frame 4 on. In consequence, our driver behavior models are able to anticipate that the passing is going to take place about .67 s before any significant, perceivable change takes place.

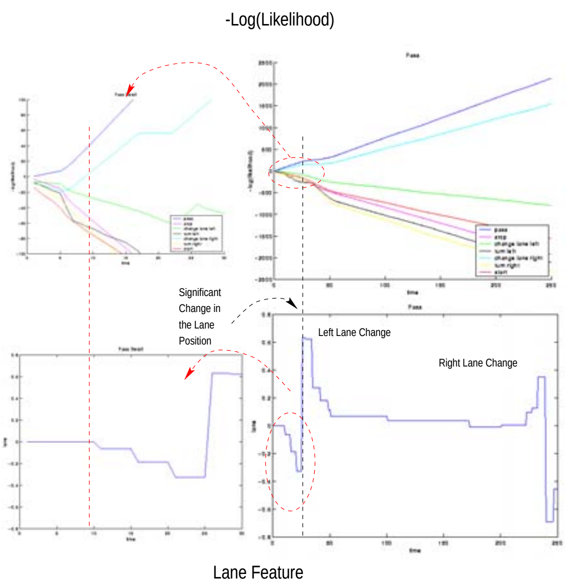

Figure 5: Prediction of a passing maneuver about 2/3 s before any significant lane change takes place.

Maneuver	Average Predictive Power in Frames (sec)
pass	37.7 (1.26 s)
stop	70.7 (2.4 s)
change lane left	2.1 (.1 s)
turn left	23.0 (.8 s)
change lane right	20.3 (.7 s)
turn right	15.1 (.5 s)
start	41.7 (1.4 s)

Table 4: Predictive power of the models in frames and seconds

6 Conclusions and Future Work

In this paper we have described our SmartCar testbed: a real-time data acquisition system in a real car and a machine learning framework for modeling and recognizing driver maneuvers at a tactical level, with special emphasis on how does the context affect the driver's performance. We have validated our theoretical framework with real driving data of over 70 subjects that drove for 1.5 hours in the greater Boston area.

We are building more complex models of driver behavior following our mathematical framework. More specifically we are building extensions of HMMs. We propose using Coupled Hidden Markov Models (CHMMs) –see figure 2– for modeling two interacting processes (humans [4] or cars). In this architecture state chains are coupled via matrices of conditional probabilities modeling causal (temporal) influences between their hidden state variables. In particular, figure 6 depicts the proposed graphical model architecture for modeling driver behavior: instead of a symmetric CHMM, we propose an asymmetric CHMM architecture, where the surrounding traffic affects the behavior of the driver, but not vice-versa. This is just an approximation to the more realistic situation of mutual interactions. The main justification of such an approximation comes from the fact that in our experiments, the driver did indeed modify his/her behavior depending on the surrounding traffic, but not vice-versa. Our preliminary results using CHMMs for

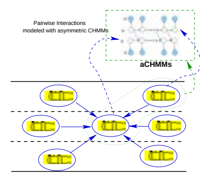

Figure 6: Representation of the asymmetric CHMMs lattice for modeling car interactions

driver behavior recognition show that: (1) the performance of CHMMs the same as that of HMMs in the worst case; (2) there are many situations in driving where a single HMM cannot capture the interactions between the driver and the surrounding traffic. It is specially in these cases when CHMMs would offer the greatest advantage. We expect to have statistics on the performance of CHMMs in the near future.

References

[1] M. Brand. Coupled hidden markov models for modeling interacting processes. *Submitted to Neural Computation*, November 1996.

[2] N. Kuge, T. Yamamura, and O. Shimoyama. A driver behavior recognition method based on a driver model framework. *Society of Automotive Engineers Publication*, 1998.

[3] N. Oliver. *Towards Perceptual Intelligence: Statistical Modeling of Human Individual and Interactive Behaviors.* PhD thesis, MIT, 2000.

[4] N. Oliver, B. Rosario, and A. Pentland. A bayesian computer vision system for modeling human interactions. In *Proc. of ICVS99*, Gran Canaria, Spain, 1999.

[5] A. Pentland and A. Liu. Modeling and prediction of human behavior. *Neural Computation*, 11:229–242, 1999.