On-Road Prediction of Driver's Intent with Multimodal Sensory Cues

By predicting a driver's maneuvers before they occur, a driver-assistance system can prepare for or avoid dangerous situations. This article describes a real-time, on-road lane-change-intent detector that can enhance driver safety.

Anup Doshi, Brendan T. Morris, and Moham M. Trivedi
University of California, San Diego

With more than 33,000 fatalities on US roads each year, and almost 50 percent of these caused by roadway departures, pervasive assistance systems with time- and safety-critical capabilities are increasingly important. The next generation of advanced driver-assistance systems (ADAS) will need to use a holistic awareness of the vehicle, its surroundings, and the driver to predict and mitigate dangerous or uncomfortable circumstances.[1]

Recent studies have demonstrated predictive systems that understand patterns of intentional driver behavior in particular situations, such as changing lanes, turning, and braking.[1,2] These systems use a wide range of vehicle sensors to develop a holistic understanding of the driving situation and machine learning algorithms to train classifiers that distinguish between patterns of events, helping to understand what the driver intends to do in the immediate future. These predictive systems could provide the early notification necessary for an ADAS to engage in assistive actions. In risky situations, the vehicle could warn the driver of impending danger or at critical moments could even take control of the vehicle to mitigate damage or completely avoid a collision. However, such a system must have an extremely low false-alarm rate if it's to be effective. A high error rate would annoy the driver, who might then disable or disregard the offending ADAS.

The implications of a system capable of recognizing and predicting the driver's intent to change lanes are numerous. Drivers only use their blinkers half the time before changing lanes,[3] but blinkers could be engaged automatically. Additionally, blind-spot warnings could be presented only when the driver needs them to avoid distraction or annoyance (to ease system acceptance). An intent-based ADAS could also provide comfort and convenience. An attempt to overtake a slower vehicle when the automatic cruise control (ACC) system is activated is more difficult than during natural driving because ACC doesn't allow throttling until after the lane change. An accurate ADAS that predicts the driver's intent to change lanes could throttle the engine earlier to mitigate the performance gap between ACC-assisted lane changes and naturalistic driving.

Analysis of intent-detection systems in recent literature has been with limited datasets using offline implementations of the systems.

Published by the IEEE CS ■ 1536-1268/11/$26.00 © 2011 IEEE

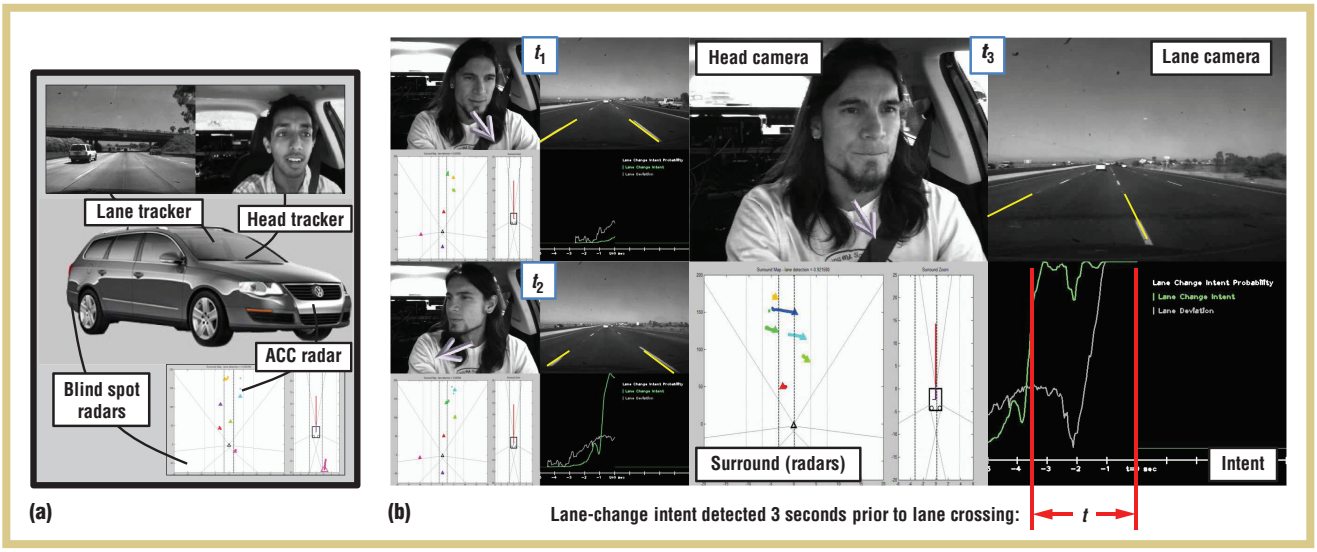


Figure 1. A lane-change prediction system developed by the Laboratory for Intelligent and Safe Automobiles. (a) LISA-X is equipped with production-grade sensors to maintain the look and feel of a stock Volkswagen Passat to ensure naturalistic vehicle operation by the driver. (b) Real-time intent prediction on LISA-X. The prediction algorithm uses the most indicative sensory signals to infer the driver's intention to change lanes.

These analyses aren't designed to determine performance in continuous, on-road implementations. We've developed and implemented a real-time lane-change-intent detection system, and we propose several analytical methods to more realistically characterize performance in general, real-world situations.

Context Capture Framework for Intelligent Vehicles

Figure 1 gives an overview of the lane-change prediction system output. We're primarily interested in answering the following questions in relation to intent-prediction systems:

- How do offline classifier performance evaluations translate to real-world, real-time, on-road performance, especially with respect to the false-positive rates (FPRs)?
- What is the performance of various sensors and sensor configurations in detecting intent? Are more advanced, costly sensors necessary to achieve better performance?
- For a given sensor configuration, when is the earliest time prior to the lane change that intent can be detected?

The Laboratory for Intelligent and Safe Automobiles (LISA) explores multidisciplinary approaches to making automobiles safer and more intelligent. We take a holistic approach to understanding the driving experience by analyzing vehicle dynamics, the vehicle surroundings, and the driver and occupants. Unlike previous LISA intelligent vehicle testbeds,[1,4] LISA-X isn't a reconfigurable research testbed. Instead, it uses production-grade sensors to understand what can be accomplished with current technology with minimal additional cost.

The LISA-X is a 2008 Volkswagen Passat Variant 3.6L, modified to include several sensors (see Figure 1a):

- ACC radar,
- side warning assist (SWA) radars,
- lane-departure warning (LDW) camera, and
- head-tracking camera (Head).

LISA-X uses the ACC and SWA radar systems to obtain obstacle information. The ACC radar uses a narrow beam to detect and track vehicles in the front, while the SWA radar system tracks vehicles in the rear. The LDW camera system tracks the lane markings on the road to determine lane-level positioning. A monocular head camera system (not yet on the market) monitors the driver's head position and orientation. We tightly integrate these sensor subsystems into the vehicle body to ensure minimal distraction during driving.

Representing a Maneuver

We outfitted LISA-X with sufficient computational resources to ensure live capture, recording, and processing of all the sensory subsystem data, as well as other signals delivered along the vehicle's controller area network (CAN) bus. During a typical drive, the system captures more than 200 sensory signals from each sensor system, synchronized and time stamped at 30 Hz, to provide a rich description of the complete driving experience, denoted by the feature vector:

$$\mathbf{x}(t) = \begin{bmatrix} \mathbf{v} \\ \boldsymbol{\alpha} \\ \boldsymbol{\varsigma} \\ \boldsymbol{\lambda} \\ \boldsymbol{\eta} \end{bmatrix} = \begin{bmatrix} \text{Vehicle} \\ \text{ACC} \\ \text{SWA} \\ \text{LDW} \\ \text{Head} \end{bmatrix}.$$

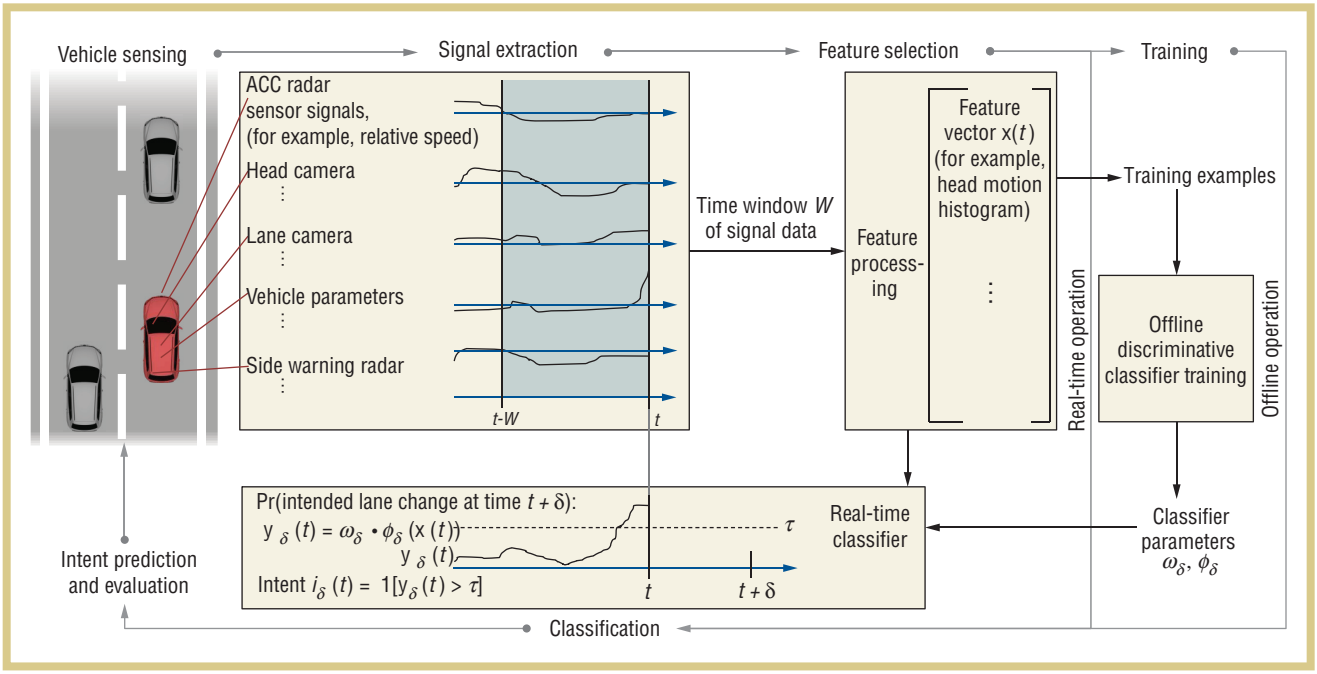


Figure 2. Overview of real-time intent-detection system. The system extracts signals from vehicle sensors and processes them into feature vectors, which it then uses for offline training and online classification.

Although each sensor subsystem provides a significant number of signals when sampled every 33 msec, the current measurements alone aren't sufficient for distinguishing between maneuvers. Instead, a small time window of the past $W = 2$ seconds worth of data is collected from each sensor to provide a temporal contextual descriptor of the driving environment (see Figure 2). The time-series descriptor incorporates temporal patterns in the $N_w = W * 30$ frames-per-second = 60 historical values. For example,

$$\boldsymbol{\eta}_{j\ldots j+N_w-1}(t) = \left[\mathrm{Yaw}(t-N_w-1),\ldots,\mathrm{Yaw}(t)\right].$$

The feature vector $\mathbf{x}(t)$ is quite large because of the N_w historical values for each of the 200+ signals. In this raw format, it would be extremely difficult for any pattern recognition algorithm to decipher without prior information about what each entry means. The raw sensor data coming is, in many cases like this, too noisy or sparse. Without enough training data for specific situations, it becomes necessary to help the system by adding "expert" knowledge, which includes extra discriminative features so classification algorithms can distinguish between different behaviors—for example, intended lane changing versus lane keeping.

The system further processes the time-series data to incorporate higher-level maneuver indicators and human driving experience, such as the amount of time the driver is glancing to the side,

$$\boldsymbol{\eta}_{j+N_w}(t) = \sum_{k=t-N_w-1}^{t} 1[\mathrm{Yaw}(k) > 30]$$

where 1[.] is the indicator function. These block features provide highly discriminative dimensions in the feature space to jump-start the machine learning process without large amounts of training data.

Subsystem Signals and Features

Each sensor type uses different feature sets to convey information. Once these signals are processed into features, they're concatenated into a large vector, as Figure 2 illustrates. Each of these vectors is used as a training example (positive or negative) for the discriminative classifier. In the case of live evaluation, we pass these vectors into the discriminative classifier to predict intent.

***Vehicle signals* (ν).** The CAN bus measures the vehicle's dynamic state and controls. This subsystem supplies several time-series features, including the steering wheel angle, yaw rate, and blinker state signals, as well as indicators of blinker direction and length of time it's active.

***ACC signals* (α).** The ACC system tracks vehicles using radar in the front of the vehicle. The ACC radar has a narrow field of view (FOV) cone such that it can reliably track only the vehicle directly in front (in the same lane, usually). The four ACC features relate to a lead vehicle and account for the distance to the ACC vehicle, the ACC vehicle's relative speed, the measured time gap with the ACC vehicle in seconds, and the difference

between the current vehicle speed and the desired speed (ACC set speed).

***SWA signals* (δ).** The SWA system tracks vehicles behind and to the sides using a paired radar system and delivers the position $(s_x^i(t), s_y^i(t))$ and relative velocity $(s_u^i(t), s_v^i(t))$ of each obstacle i. Unlike the ACC radar, the SWA system can track many vehicles simultaneously because it has a much larger FOV. The large rear area is quantized into three smaller critical zones corresponding to the blind spots in the rear of the vehicle, as Figure 3 shows. Each zone is between $-15 < y < -5$ meters behind and is defined by the size of adjacent lanes: $Z_1 = \{x \mid -5 < x < -1.65\}\}$; $Z_2 = \{x \mid -1.65 < x < 1.65\}\}$; and $Z_3 = \{x \mid 1.65 < x < 5\}\}$.

The SWA blind-spot features indicate the occupancy and speed state within a critical zone z as

$$\varsigma_z = \max_i \frac{1}{N_w} \sum_{k=0}^{N_w-1} s_v^i(t-k),$$

where i iterates through all tracked vehicles at the current time t. The features ς_z represent motivating or deterring factors for a lane change because the presence of a vehicle in the adjacent lane would impede the lane change.

***LDW signals* (λ).** The LDW system measures the vehicle position with respect to the road and the road geometry. The LDW features correspond to the recent time series of vehicle lateral deviation (position within the lane), lane curvature, and vehicle yaw angle with respect to the lane.

***Head signals* (η).** Unlike the other sensor subsystems, the Head system monitors the driver instead of the driving environment. A driver's intentions can be better inferred when directly measuring driver actions. The Head features are a measure of recent driver head motion, both head rotation (yaw) and pitch. The features include the time series of head yaw position, yaw motion (derivative of yaw position), a histogram of head yaw values, a histogram of yaw motion values, a histogram of head pitch position, and an indicator signal of significant yaw rotation. Because preparatory glances are a major indicator of a lane-change maneuver,[3] many features are generated for the Head system.

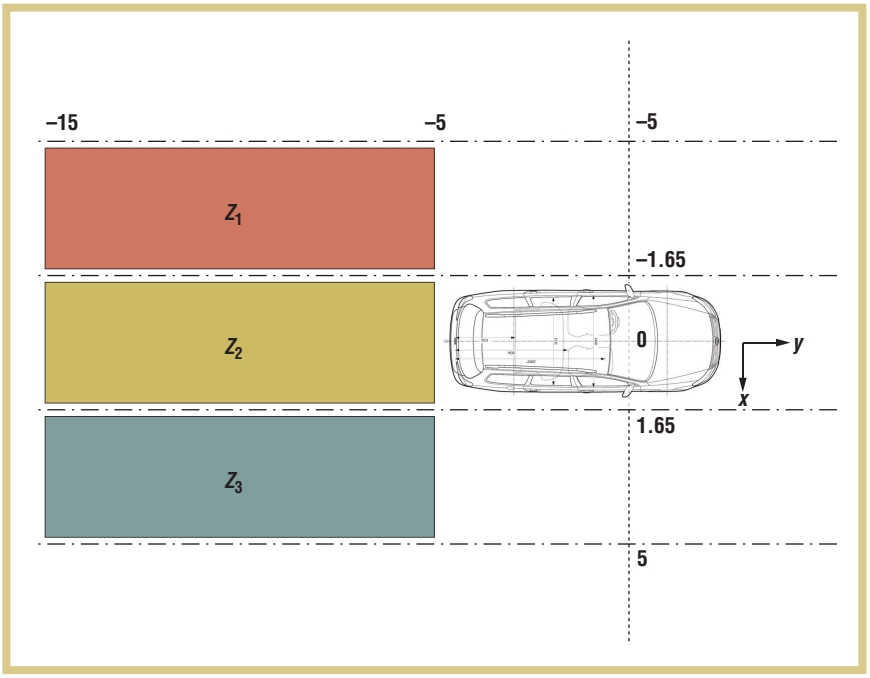


Figure 3. Three side warning assist (SWA) regions of interest: driver-side lane in pink, center lane (or *ego-vehicle lane*) in tan, and passenger-side lane in blue. The SWA features consist of the average longitudinal speed in each of these regions and indicate motivators and deterrents to a lane change.

On-Road Intent-Prediction Framework

To train and test the lane-change-intent classifier, we collected a new sensor-rich database of naturalistic driving using the instrumented LISA-X vehicle. We collected the data in Palo Alto and San Diego in conjunction with and under the supervision of the VW Electronic Research Laboratory (ERL). The 15 (12 male, three female) participants were of various nationalities and ranged from 20 to 50 years old, with corresponding amounts of driving experience—from several years to decades. Familiarity with ACC functionality ranged from very little to advanced everyday users. Especially for those drivers with limited experience, we provided a brief ACC tutorial that included sample driving time until they became comfortable with operating the system.

For each driver, a data collection run consisted of two phases designed to capture a variety of behaviors. In one phase, we told the driver to operate with the ACC function active. After a brief tutorial and once the driver was comfortable with ACC, we asked the driver to safely engage in highway driving. We told the driver to stay in the slow lane and engage ACC when it was safe and comfortable to do so, passing other cars when necessary (lane changes could occur on either the driver or passenger side).

In the second collection phase, in order to obtain completely natural driving behaviors, we allowed the driver to drive normally without the ACC function engaged. Drivers (especially those not familiar with ACC operation) tended to behave differently when the

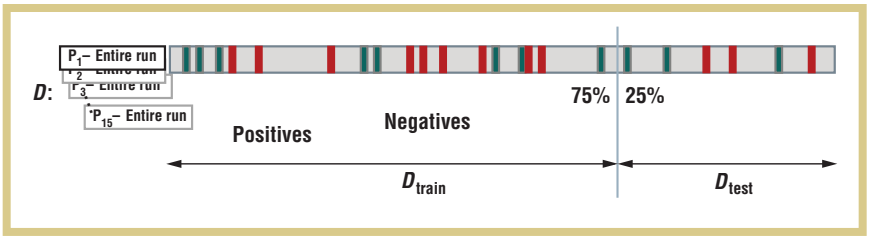


Figure 4. Training database split. The training database, which consisted of 15 subjects, was split into a training and cross-validation test set to train the lane-change classifier. We randomly chose 75 percent of the lane changes from each subject to use for training. The remaining 25 percent was for independent testing. Each positive or negative example corresponds to a small 2-second segment of driving data.

TABLE 1
Lane-change prediction datasets.

Label	Dataset	N_{people}	N_{runs}	Hours	Description
D	Long runs for training	15	24	14.5	782,189 frames
D_{train}	Training examples				266 positive/2,606 negative examples
D_{test}	Test examples				101 positive/879 negative examples
T	Long runs for testing	4	18	7.9	427,497 frames

ACC was active than in their normal, day-to-day driving.

The two phases lasted approximately one hour in total, with 30 minutes dedicated to each collection phase, letting us capture sufficient data to find patterns in various conditions. The complete training dataset D contained 24 driving runs for 14.5 total hours of driving data (CAN messages and raw videos).

All the lane changes in a collection run were automatically detected using the lane-tracking information. We can detect lane changes by observing the point at which the lateral lane position signal switches sign (from the right to left side of the lane, or vice versa). The automatic-detection scheme worked well when lane tracking was reliable, but several lane changes had to be manually marked by examining forward-looking video corresponding to those periods of broken lane detection.

The automatic lane-change-detection scheme identified almost 500 examples (with a 2-second long data window W) that could be extracted as training segments. To ensure that the training data was valid, we used several criteria to discard any data in which the sensors might have performed poorly:

- Vehicle: the ego-vehicle must travel at highway speeds.
- ACC: the cruise control must be on (but not necessarily active).
- Head: the visual head tracking must be consistent by having limited abnormal head movements and the driver's head located within the center of the camera FOV. There can't be too many discontinuities in the head rotation time series.
- LDW: the lane tracking should be consistent (with high confidence) over the data window.
- SWA: none.

There were approximately 370 instances of valid naturalistic lane changes in the 15 driving hours from dataset D. We split the training data into two separate subsets for cross-validation training. The set D_{train}, with 266 positive examples, was for training the classifier, whereas D_{test}, with 101 positive examples, was for assessing the performance during training, with approximately 10 times as many negative examples in each case. Figure 4 depicts the data split methodology.

Our testing dataset consisted of a subset of our members of the initial training data and was collected separately from the training data. Unlike during training, the testing dataset T consisted of 7.9 hours of completely uncontrolled driving in 18 separate data runs. Drivers were free to use ACC as desired and weren't prompted with any directions. Dataset T contained more than 400,000 evaluation windows but only 229 lane changes. Table 1 summarizes the lane-change prediction datasets.

We based our lane-change-intent inference algorithm on the Driver Intent Inference System (DIIS), by Joel McCall and his colleagues.[2] There are a number of other works in this field.[5–9] Our experience developing discriminative classifiers for intent prediction has shown considerable success, warranting this investigation into a real-time prototype.[1,2]

The DIIS is a discriminative classifier that distinguishes between lane changing (either right or left) and lane keeping. Figure 2 shows the general flow of the system.

To train the DIIS, we used the discriminative relevance vector machine (RVM) classifier, which is based on sparse Bayesian learning (SBL), developed by Michael Tipping[10] and implemented by McCall and his colleagues.[2] The algorithm is a Bayesian counterpart to the popular support vector machines (SVM) and is used to train a classifier that translates a given feature vector into a class membership probability. RVMs use a parameterized prior to prune large feature vectors and facilitate a sparse data representation.

The basic form of the RVM classifier is $y_\delta(t) = \omega_\delta \times \varphi_\delta(x(t))$, where x is the input feature vector, ω_δ is the

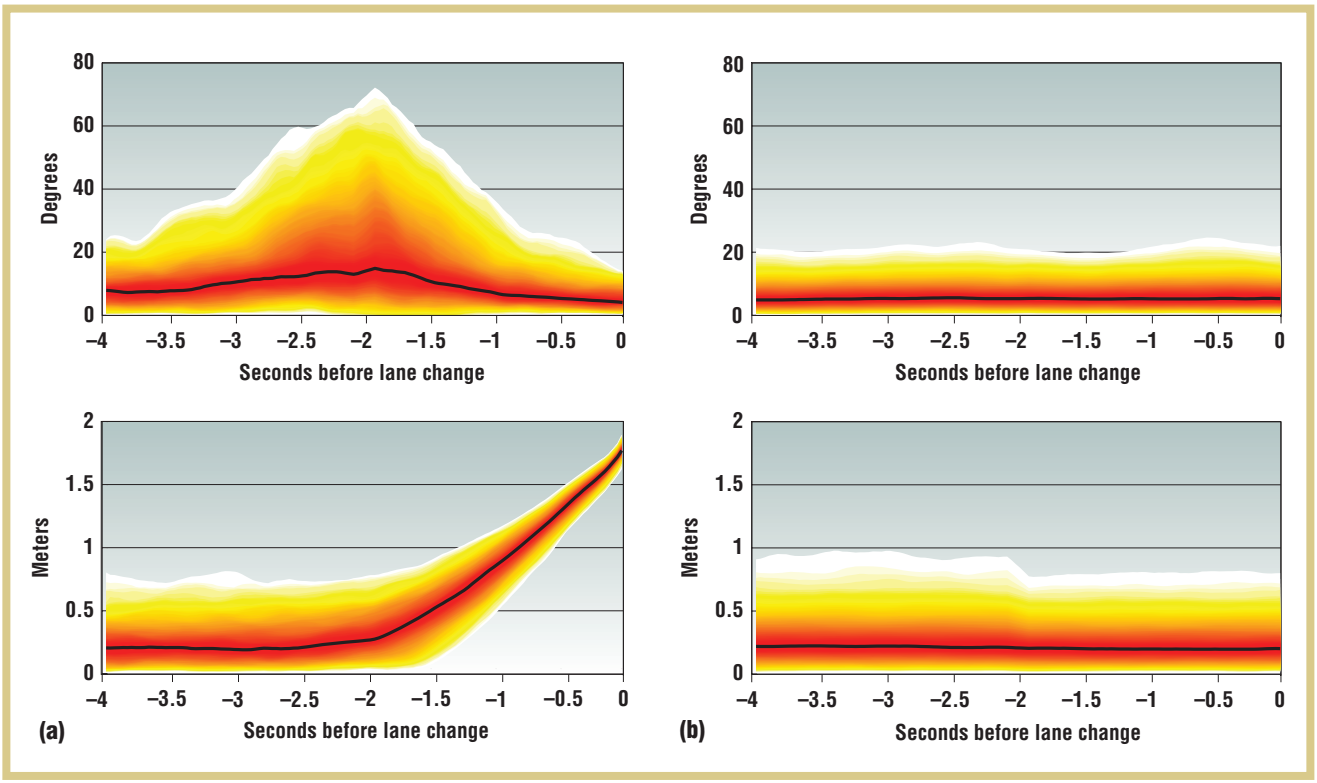


Figure 5. Comparison of driver features. (a) Evolution of head yaw (top) and lateral lane position (bottom) before a lane change when lane deviation is maximal. A driver's head yaw (rotation) when scanning tends to occur between 2.5 and 1.5 seconds before a lane change, much later than was previously reported.[3] (b) Typical movements during normal, nonlane change events.

learned model weight, and φ_δ is a kernel function. The output y represents the probability that x belongs to a particular class. In this case, we determine whether x represents an intended lane change at a particular time δ in the future.

Several advantages of this methodology motivate the use of RVMs over other algorithms, such as SVMs and hidden Markov models (HMMs). The RVM can sift through large feature sets and obtain a sparse data representation, which is especially useful in this application in identifying a small set of useful features (those that distinctly precede a lane change, as Figure 5 demonstrates). Multimodal data from various sets of sensors can thereby be combined easily, with the RVM automatically choosing discriminating cues from each modality. The resulting sparse representation allows for quick computation and classification in real time and real-world conditions with limited hardware.

The SBL methodology is general enough to consider cases, such as in our experiment, where there are relatively few training examples as compared with the number of features. By including the windowed time series of cues in the final feature vector and applying the kernel function, the RVM can also determine nonlinear temporal relationships between features, eliminating the need for HMMs.

On-Road Performance Characterization

The performance of pattern-recognition-based classifiers in laboratories and real-world use differ significantly. In the laboratory setting, prespecified examples are fed to the classifier to determine performance; in real-world settings, however, the classifier must evaluate times-series data as it arrives. This continuous operation directly impacts performance because the amount of data to be evaluated is much greater, consecutive evaluations will perform similarly, and decisions must be made in real time. Practically, this results in much lower FPRs because of the significantly larger number of evaluations (although this doesn't necessarily translate to "better" performance) and necessitates the development of techniques to perform consistent labeling of consecutive samples.

From Laboratory to Roads

During the intent classifier training, we evaluate performance using a small cross-validation test set D_{test}. The receiver operating characteristic (ROC) curve generated using this set of labeled positive and negative examples (Figure 6a) provides the traditional performance evaluation in the laboratory setting and is similar to those seen in prior work.[2]

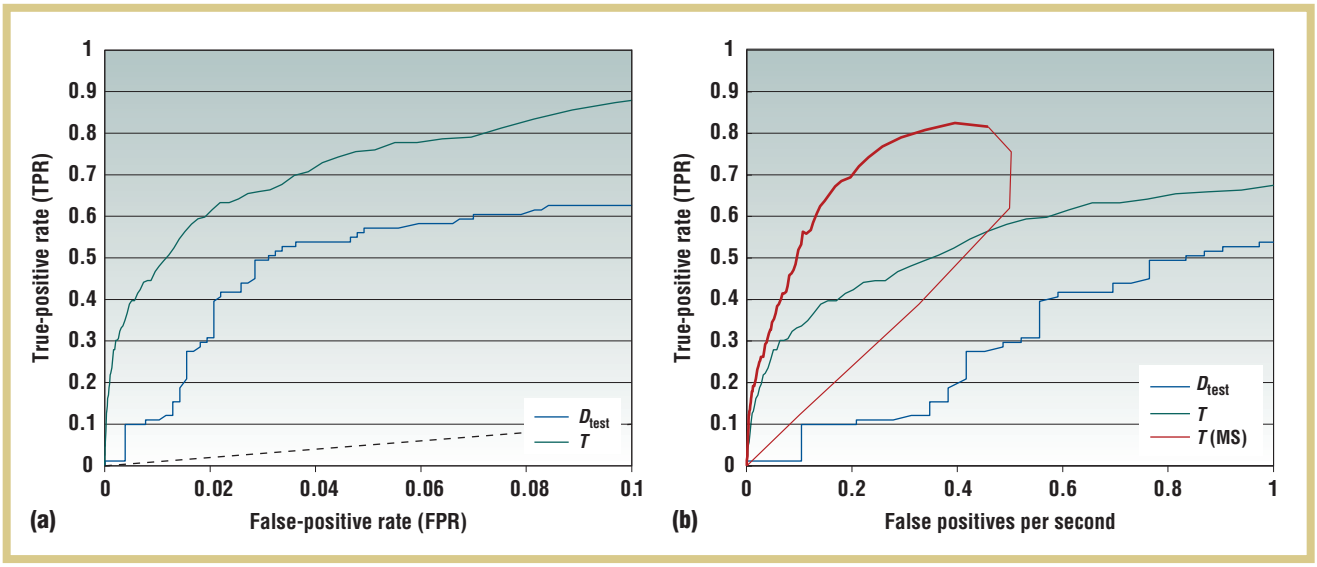


Figure 6. Plots comparing the $\delta = 2.5$ intent receiver operating characteristic (ROC) curves with all sensor subsystems for the different test sets. (a) The traditional true-positive rate (TPR) versus false-positive rate (FPR) ROC curve used to assess classifier performance with results on *T* higher because of the many more negative examples in the set. (b) An ROC curve using real units shows that the laboratory evaluation doesn't translate well to the road. Lowering the number of false positives using a multisuppression (MS) technique can greatly improve the on-road performance, as shown in red.

As one of the design requirements, we aimed to create an algorithm that can predict behaviors across a wide population of drivers. This includes drivers of different styles, some of whom might engage in slightly different behaviors prior to changing lanes than others. Given the variety of styles, we analyzed the classifier's performance on individual drivers to understand whether performance degrades between drivers. We compared the class membership probabilities between the positively and negatively labeled data (that is, lane changes versus lane-keeping situations). Using an analysis of variance, treating the subjects as random factors, we found a consistent pattern across all drivers of separability between classes ($F(1, 14) = 674$; $p < 0.001$). Across the range of drivers, the classifier performs consistently, and we expect it to perform just as well on a larger population.

We then moved onto dataset *T*, which had fewer drivers but much longer and continuous drives, to analyze the on-road performance of the predictor response during continuous operation. Figure 6a shows the two ROC curves (in blue and green) for detection time $\delta = 2.5$ using all the sensor subsystems. The ROC from the on-road set *T* is better than for D_{test} because of the greater rate of negative examples, resulting in a low FPR.

Although the traditional ROC curve analysis makes the classifier seem quite strong, viewing the results on a scale that's more meaningful for real-world usage reveals a different story. Figure 6b presents the same ROC but using false positives per second rather than the traditional FPR. Rather than use academic measures of performance, this lets us interpret concrete meaning in terms of real units. For example, a classifier with a 70 percent true-positive rate (TPR) and 5 percent FPR might seem reasonable in an academic sense, but during real on-road operation, it might translate to more than one false positive every second. Therefore, good classifier performance in the laboratory doesn't necessarily translate well on-road. To ensure good on-road performance, we must greatly improve the classifier FPR.

Lowering Classifier FPR

The intent classifier performance analysis using real units (FP/sec) shows that false detections are the primary source of error. These false positives are mainly the result of two phenomena. First, on-road evaluation presents data sequentially and continually, which results in nonindependent outputs. Second, a classifier is designed to detect a lane change δ seconds before its occurrence, but the classifier isn't always so precise in its prediction.

Time-series dependency and multisuppression. In classical pattern-classification setups, data is selected (either positive or negative examples) and fed to a classifier to test its response. The implicit assumption is that individual examples are independent. In the on-road case, we use a sliding window of data on the time series of feature data, resulting in nonindependent samples. Two consecutive windows will contain almost identical data and therefore will have similar responses. The plot in Figure 7 shows the

evolution of intent probability of the $\delta = 2.5$ classifier in black for several seconds before a lane change.

Ideally, a single large value would spike at 2.5 seconds prior to the lane change, but here there are actually two major responses, at 3.5 and 1.25 seconds prior. As the lane change approaches, the probability of lane change increases and a number of time instances are above the detection threshold. Each red delta indicates a detection, but there is only one lane change, which means most red deltas are false positives. Assuming that consecutive detections arise from the same intent, we can extract a single detection on the rising edge of the intent signal. This multiple detection-suppression technique logically handles the time-series effect but might change the detection time away from δ. After this multisuppression, only three detections remain in this segment. The detection that matches the lane change is the closest to δ, as shown in green, whereas the remaining two detections are false detections—dramatically fewer than the red.

The red line in Figure 6b shows the on-road performance after multisuppression. The multisuppression scheme results in performance gain over traditional evaluation but introduces an interesting looping behavior in the ROC curve. At low-threshold values, many samples are considered continuous detections and then suppressed. Therefore, the false-detection rate will be low but true detections are also low, because the multisuppression response won't match the desired detection time well. Because this effect occurs only at low thresholds, it is not much of a concern in practice; the selected detection threshold would be high to have high confidence in the prediction and to limit false positives.

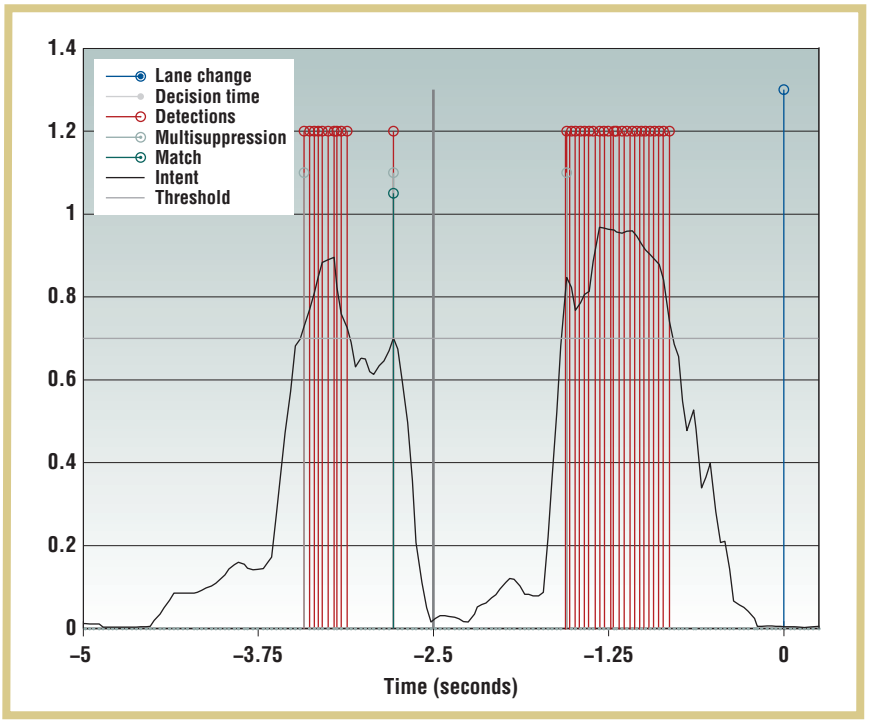


Figure 7. Multisuppression. The blue line indicates the time of a lane change, and the red lines denote all of the times when the prediction was above the threshold. Because only one lane change occurs, a single match, in green, is a true positive. The remaining red lines are false positives. We can manage the number of false positives by suppressing multiple hits and only counting a single response from each of the two groups of red lines (as shown in light gray).

Imprecise prediction timing and match windows. Although we trained an intent classifier to infer the lane change δ seconds before its occurrence, it's merely attempting to match the current data to a pattern seen during training. The cues that signify an oncoming lane change might not always happen at the exact same time, due to the variability in human behaviors. Figure 8 plots the timing response of an intent classifier for different detection times. To generate these plots, we found all time instances where the prediction probability value was above a low detection threshold and determined the time that elapsed before the next lane change. Although a detector is designed for a specific time δ, the detections aren't localized and can occur before or after this time. Therefore, it isn't sufficient to only consider the sample δ seconds before the lane change as a positive example, we must also consider that examples within a small time window around δ are positive.

To determine if the classifier was able to correctly infer the lane change δ seconds early, we used a small match window to soften the δ constraint to consider a close detection as a match (the green line in Figure 7). The match window size is parameterized by two values, α and f,

$$t^{+}_{mw} = f\delta$$
$$t^{-}_{mw} = \alpha f\delta$$

as shown in Figure 9. The match window is not symmetric to bias matching for earlier detection. The early bias is a result of multisuppression, which tends to mark one detection before the maximal response. The f term characterizes the timing of detection spread, whereas α enables earlier detections. We use $\alpha = 2.0$ for all further analysis because performance improvements are limited for larger values.

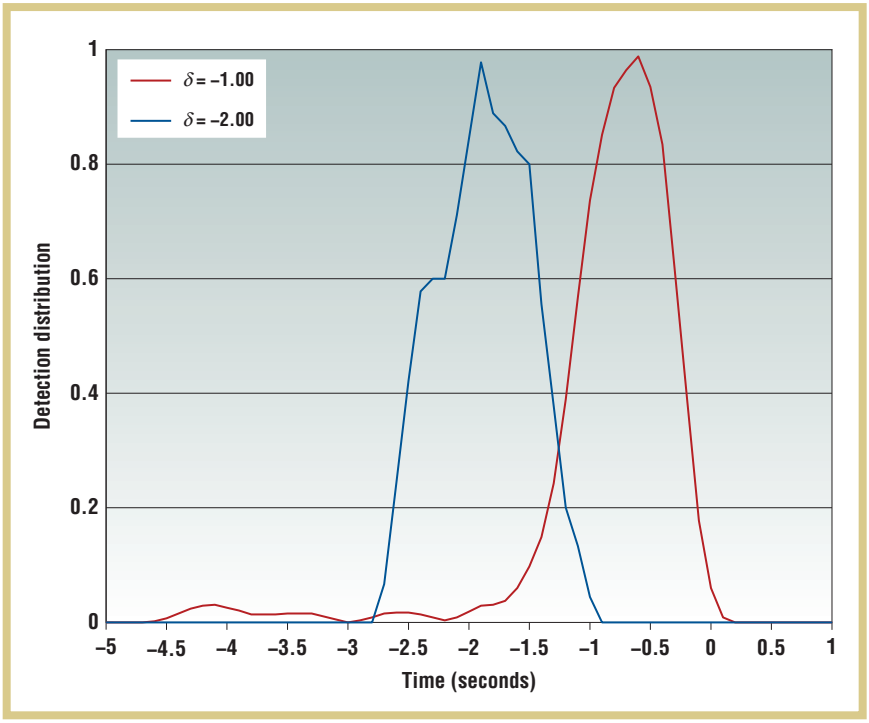


Figure 8. The distribution of intent detections is fairly wide because consecutive intent classifications will result in a succession of positive results. A positive detection doesn't always occur exactly at detection time but can occur sometime before and after.

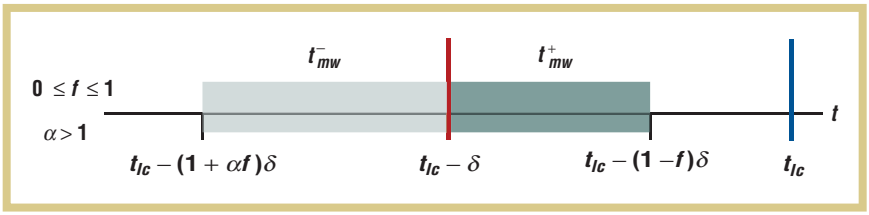


Figure 9. The match window's size is controlled by the filter detection time δ and the two parameters, f and α. Parameter f characterizes detection spread and α enables earlier detections.

Figure 10a shows the detection rates for intent classifiers with different detection times increasing with larger match window sizes. Of course, larger windows will more likely collect a high response, but the cost is an increased number of false positives within the window, as Figure 10b shows. The FPR dips after initially increasing because the number of negative examples in a window grows faster than the number of new false positives. This peak gives a good indication of the natural window size for a particular detection time δ.

Selection and Performance of Different Sensor Configurations

Each sensor that's added to an automobile provides both new functionality and a more complete understanding of the driving situation. However, these sensors only provide measurements of the physical world, and it isn't immediately clear when they're most useful for lane-change-intent recognition.

Subsystem Timing Exploration

It's important to know at what time period before a lane change each of the cues from a particular subsystem are relevant. We can use this information to design a classifier that uses the sensors appropriately.

Figure 11 shows the performance of each individual sensor subsystem over a range of detection times. The color indicates how important a sensor system is compared to other sensors using the area under the ROC curve (AUC) criterion. The LDW lane information is most informative between 0 and 1.5 seconds prior to a lane change, as Figure 5 demonstrates. The Head information is most relevant between 2 and 3 seconds, corroborated by the histogram of raw head movements in Figure 5. Interestingly, the SWA system doesn't seem to add much information through 5 seconds, which could indicate that the features aren't informative or that this system can't be used alone. The ACC radar apparently is the most informative sensor at 5 seconds. By training intent classifiers that only use measurements from a particular sensor system, we can determine its relevant time horizon without deep analysis or understanding of the signals themselves.

Sensor Configuration Exploration

It's reasonable to assume that having more sensors will provide better information. Although more data might provide a more complete picture of the driving situation, it's also more difficult to manage and effectively use all the information (aka the "curse of dimensionality"). In addition, these automotive sensors aren't standard equipment on vehicles but are added options, which increase the vehicle's cost. Therefore, we want to know the relative performance of different sensor configurations and whether all the advanced sensor systems are necessary to accurately predict intent.

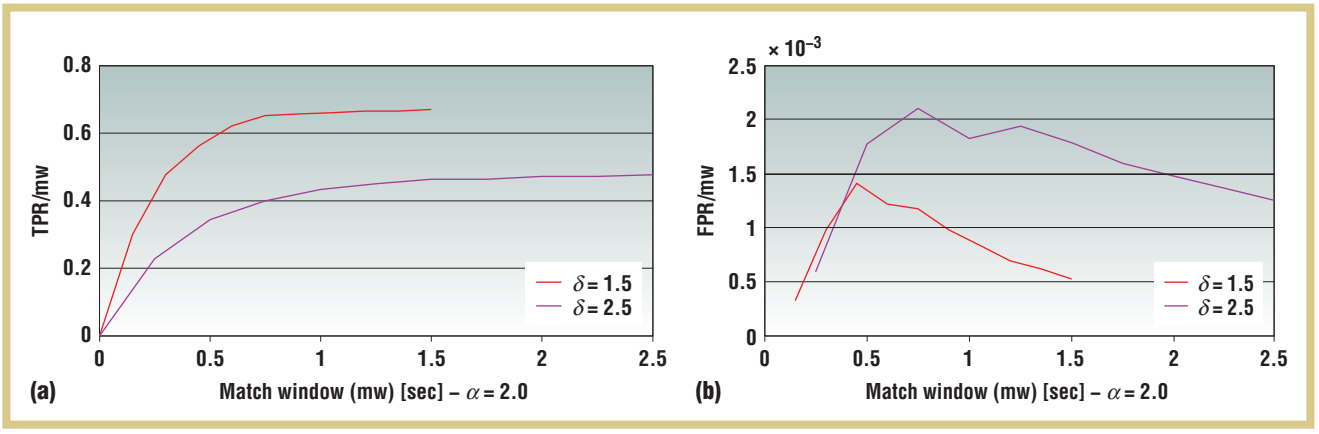


Figure 10. Detection rate. (a) The true-positive rate (TPR) increases for different match window sizes. The strong classifiers can increase rapidly with small window increase (between 0 and 1 second). (b) The cost of better detection is a larger number of false positives within the match window.

We examined several sensor configurations corresponding to different equipment options. For lane-change detection, we require at least the vehicle information and the LDW lane information. We evaluated the following sensor combinations:

- VLAHS (vehicle, LDW, ACC, Head, SWA),
- VLAH_ (vehicle, LDW, ACC, Head),
- VL_HS (vehicle, LDW, Head, SWA),
- VL_H_ (vehicle, LDW, Head),
- VLA_S (vehicle, LDW, ACC, SWA),
- VLA_ _ (vehicle, LDW, ACC), and
- VL_ _S (vehicle, LDW, SWA).

All previously presented results and discussion referred to the full sensor combination (VLAHS).

Figure 12 compares the real-unit ROC curves for detection times $\delta = 1.0$, 1.5, and 2.5 seconds. For the $\delta = 1.5$ classifier in Figure 12a, the detection rate for all the classifiers can reach 95 percent after 200 FP/hour. The VLA_ _ and VL_ _S perform the worst. At this time, the use of all the sensors in VLAHS actually degrades performance slightly. Figure 12b shows a clear separation between the worst classifier, VL_ _S, the non-Head classifiers (VLA_S and VLA_ _), and the Head-based classifiers. At $\delta = 1.5$, we can see improvement in performance through use of the head-viewing camera. Similarly, at $\delta = 2.5$ the Head-based predictors are the best performing, but this time the SWA only (VL_ _S) classifier outperforms ACC-based systems. These results are consistent with prior research that found direct observation of the driver improved lane-change-intent prediction by detecting preparatory scanning.

As expected, the intent classification quality improves the closer the detection time is to the lane change.

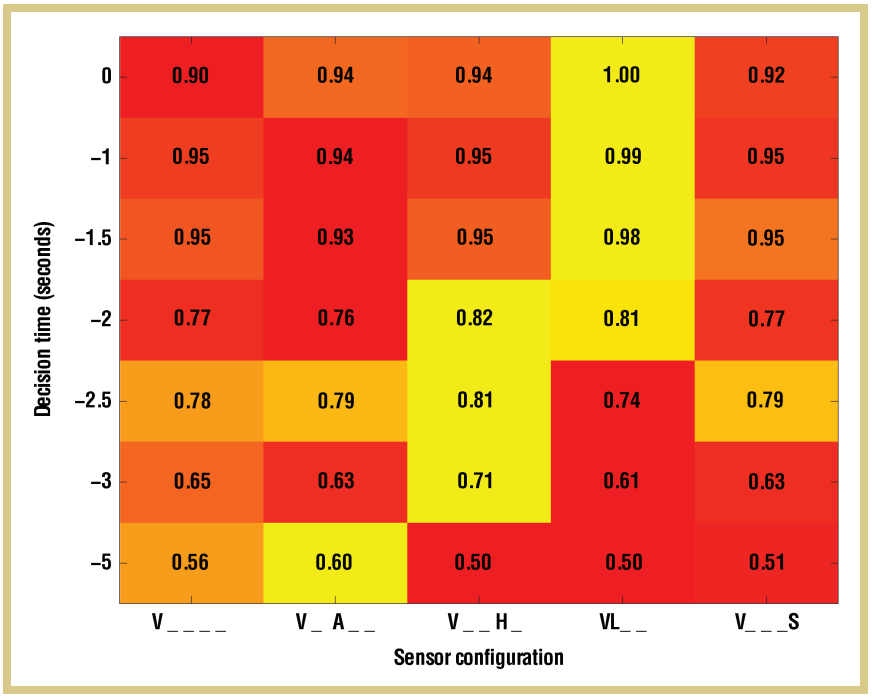


Figure 11. Different sensor subsystems contribute relevant information at different times before a lane change. The color indicates how important a sensor system is at a given decision time compared to the others using the area under the receiver operating characteristic (ROC) curve (AUC).

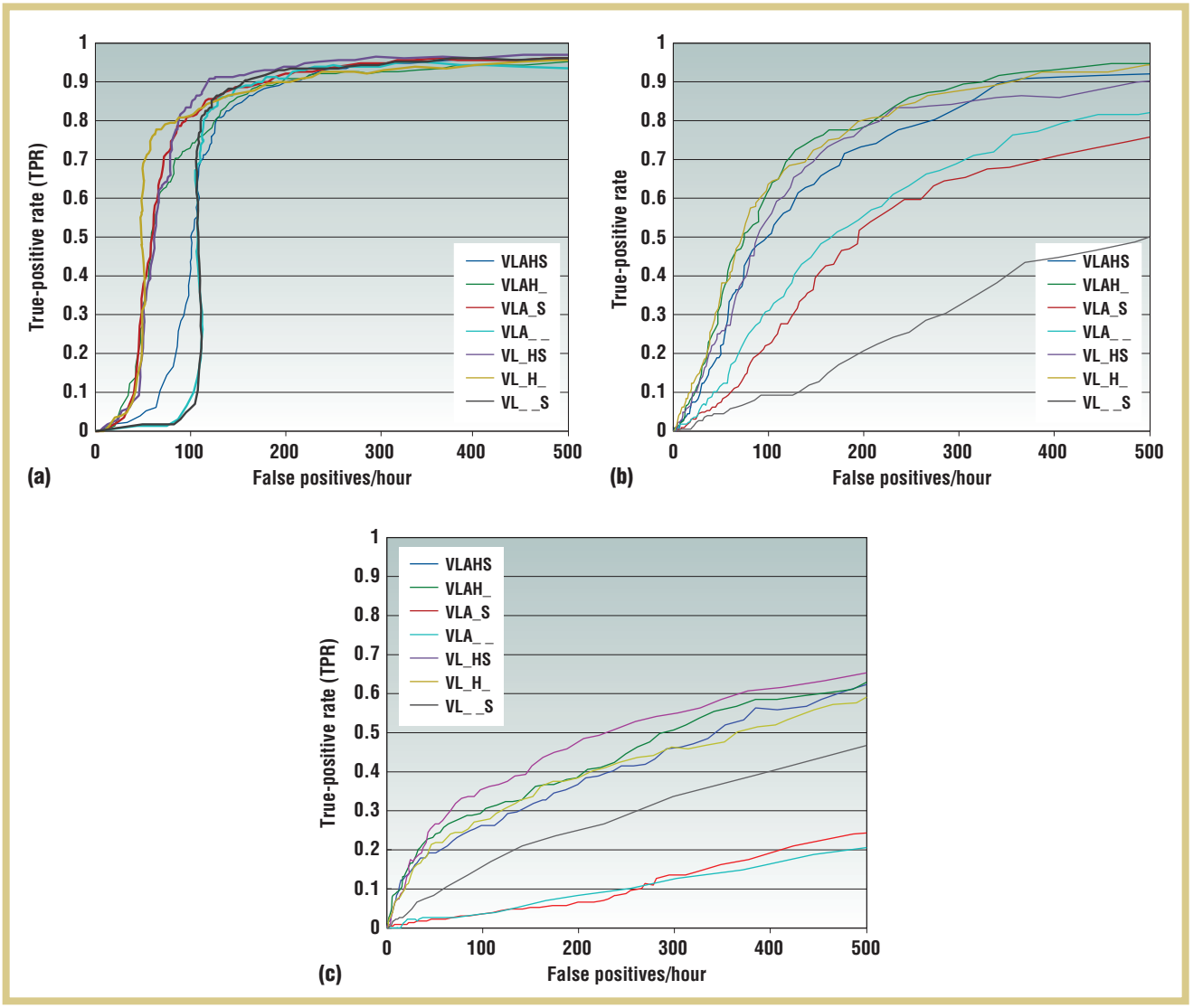


Figure 12. Real-unit receiver operating characteristic (ROC) curves for competing sensor configurations at different detection timings: (a) $\delta = 1.00$, (b) $\delta = 1.50$, and (c) $\delta = 2.50$.

Figure 13 shows strong performance up to $\delta = 1.5$, and that intent prediction seems to be unreliable beyond $\delta = 3.0$ seconds. This indicates that intention manifests itself in movement only in the 3 seconds leading up to the lane change.

For $\delta = 0.0$, the ROC curve bends toward the top left corner during the increasing TPR. This is an artifact of the multisuppression algorithm not normally present in ROC plots. At high thresholds, the multisuppression algorithm can't conglomerate detections because of noisy peaks. Once the threshold falls below this noise level, more consecutive frames are merged through the multisuppression procedure, reducing the number of false positives.

Finally, Figure 14 presents a complete comparison between all the sensor configurations and detection timings. This matrix indicates the detection rate for a fixed FPR of 120 FP/hour. Generally speaking, the performance for all sensor configurations decreases further away from the lane change except for VL_ _S (local maximum at $\delta = 2.5$). The color scale indicates the relative strength of a sensor configuration at a given detection time. The coloring shows that the sensor configurations using the Head sensor have consistent gains between $\delta = 1.5$ and 2.5 seconds, making it an attractive advanced sensor.

Automotive systems operate under special safety and time-critical constraints, where every millisecond

helps to save lives. Intelligent driver-assistance systems, along with more general pervasive computing systems, gain precious moments and thus stand to benefit greatly from proactively understanding and predicting human behaviors.

ACKNOWLEDGMENTS

A grant from the University of California Discovery's Digital Media Innovation Program and the Electronics Research Laboratory of Volkswagen of America made this research possible. We appreciate the advice and assistance of our reviewers along with our LISA and ERL colleagues, and specifically the contributions of Jaime Camhi at ERL.

REFERENCES

1. S.Y. Cheng and M.M. Trivedi, "Turn Intent Analysis Using Body-Pose for Intelligent Driver Assistance," *IEEE Pervasive Computing*, vol. 5, no. 4, 2006, pp. 28–37.
2. J.C. McCall et al., "Lane Change Intent Analysis Using Robust Operators and Sparse Bayesian Learning," *IEEE Trans. Intelligent Transportation Systems*, vol. 8, no. 3, Sept. 2007, pp. 431–440.
3. S.E. Lee, C.B. Olsen, and W.W. Wierwille, *A Comprehensive Examination of Naturalistic Lane Changes*, tech. report DOT HS 809702, Nat'l Highway Traffic Safety Administration (NHTSA), US Dept. of Transportation, 2004.
4. M.M. Trivedi, T. Gandhi, and J. McCall, "Looking-In and Looking-Out of a Vehicle: Computer-Vision-Based Enhanced Vehicle Safety," *IEEE Trans. Intelligent Transportation Systems*, vol. 8, no. 1, Mar. 2007, pp. 108–120.
5. N. Kuge et al., "A Driver Behavior Recognition Method Based on a Driver Model Framework," *Proc. Soc. Automotive Engineers World Congress*, Soc. of Automotive Engineers, 2000.
6. N. Oliver and A.P. Pentland, "Graphical Models for Driver Behavior Recognition in a Smartcar," *IEEE Proc. Symp. Intelligent Vehicles*, IEEE Press, 2000, pp. 7–12.
7. D.D. Salvucci et al., "Lane-Change Detection Using a Computational Driver Model," *Human Factors*, vol. 49, no. 3, June 2007, pp. 532–542.

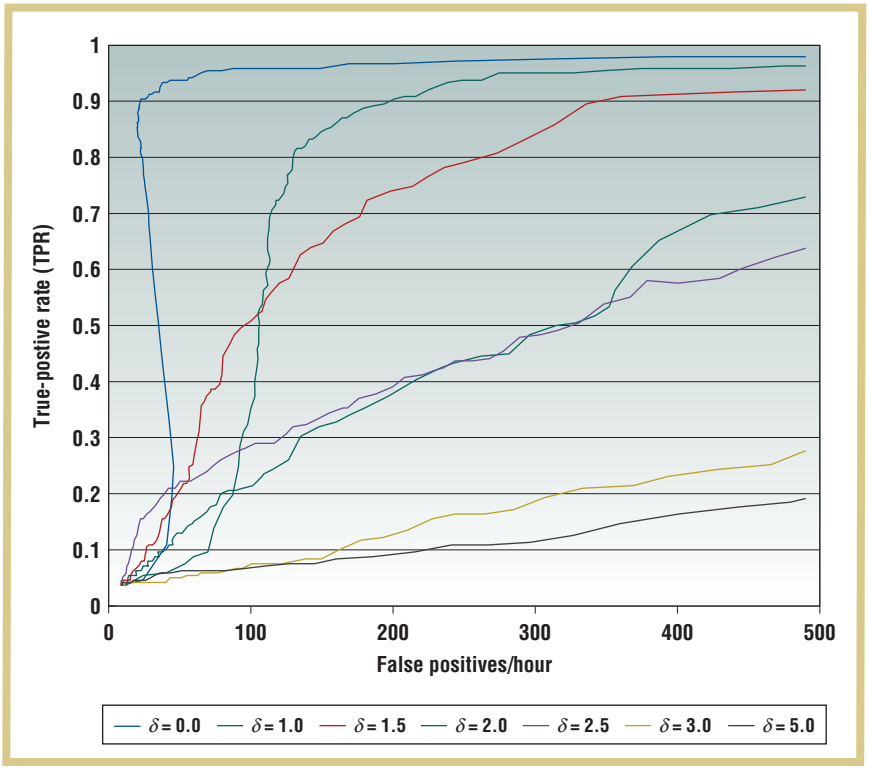


Figure 13. Real-unit receiver operating characteristic (ROC) curve for the VLAHS intent classifier at different detection timings. The closer to the lane change, the better the performance. Beyond 3 seconds, prediction is completely unreliable.

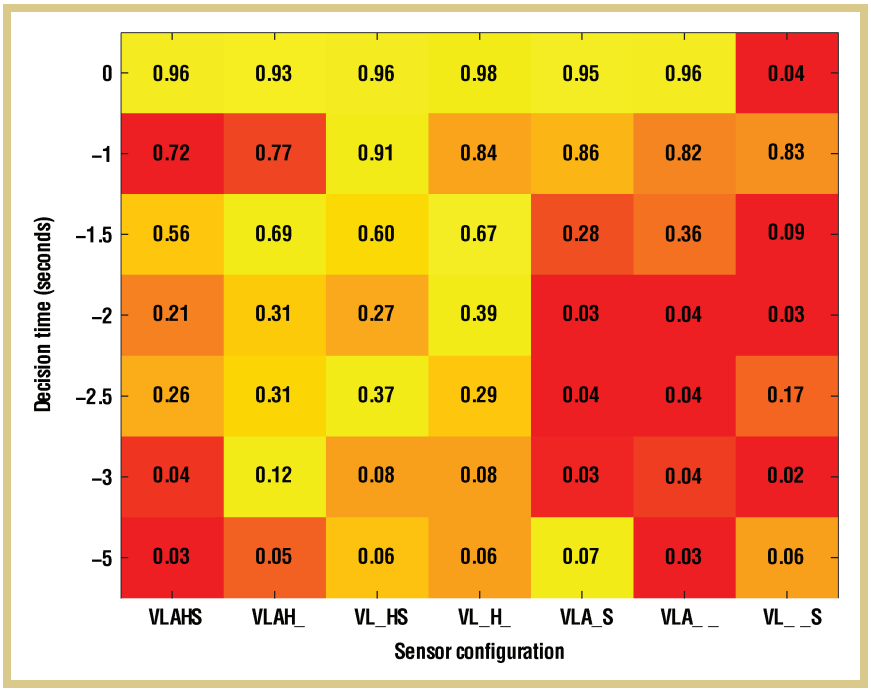
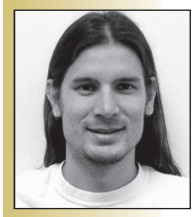


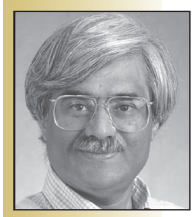
Figure 14. Intent detection rate for a fixed 120 false-positive per-hour rate (FP/hour). Similar to the individual sensor study, using the head sensor to observe the driver actions improves the lane-change intent between 1.5 and 2.5 seconds.

the AUTHORS

Anup Doshi is a researcher in the Laboratory for Intelligent and Safe Automobiles (LISA) at the University of California, San Diego. His research interests include computer vision and machine learning for applications in driver assistance and intelligent vehicles, human-computer interfaces, and behavior prediction. Doshi has a PhD in electrical engineering (signal and image processing) from UCSD. He is a member of IEEE. Contact him at andoshi@ucsd.edu.



Brendan T. Morris is a postdoctoral researcher in the Computer Vision and Robotics Research Laboratory at the University of California, San Diego. His research interests include intelligent vehicles, intelligent transportation systems, and distributed vision and information archival and visualization. Morris has a PhD in electrical and computer engineering from UCSD. He is a member of IEEE. Contact him at b1morris@ucsd.edu.



Mohan Manubhai Trivedi is a professor of electrical and computer engineering and the founding director of the Computer Vision and Robotics Research Laboratory at the University of California, San Diego. His research interests include intelligent systems, computer vision, intelligent environments, intelligent vehicles and transportation systems, pervasive computing, and human-machine interfaces areas. Trivedi has a PhD in electrical engineering from Utah State University. He is a fellow of IEEE, the International Association for Pattern Recognition (IAPR), and SPIE. Contact him at mtrivedi@ucsd.edu.

8. M. Itoh, K. Yoshimura, and T. Inagaki, "Inference of Large Truck Driver's Intent to Change Lanes to Pass a Lead Vehicle via Analyses of Driver's Eye Glance Behavior in the Real World," *Proc. SICE Ann. Conf.*, IEEE Press, 2007, pp. 2385–2389.

9. M.J. Henning, O. Georgeon, and J.F. Krems, "The Quality of Behavioral and Environmental Indicators Used to Infer the Intention to Change Lanes," *Proc. 4th Int'l Driving Symp. Human Factors in Driver Assessment, Training, and Vehicle Design*, Univ. of Iowa, 2007, pp. 231–237.

10. M.E. Tipping and A.C. Faul, "Fast Marginal Likelihood Maximization for Sparse Bayesian Models," *Proc. 9th Int'l Workshop on Artificial Intelligence and Statistics*, Soc. for Artificial Intelligence and Statistics, 2003.

cn Selected CS articles and columns are also available for free at http://ComputingNow.computer.org.