Journal of
Electronic Imaging

SPIEDigitalLibrary.org/jei

# Driver hand activity analysis in naturalistic driving studies: challenges, algorithms, and experimental studies

Eshed Ohn-Bar
Sujitha Martin
Mohan Manubhai Trivedi

# Driver hand activity analysis in naturalistic driving studies: challenges, algorithms, and experimental studies

**Eshed Ohn-Bar**
**Sujitha Martin**
**Mohan Manubhai Trivedi**
University of California
Laboratory of Intelligent and Safe Automobiles
San Diego, La Jolla, California 92093
E-mail: eohnbar@ucsd.edu

**Abstract.** *We focus on vision-based hand activity analysis in the vehicular domain. The study is motivated by the overarching goal of understanding driver behavior, in particular as it relates to attentiveness and risk. First, the unique advantages and challenges for a nonintrusive, vision-based solution are reviewed. Next, two approaches for hand activity analysis, one relying on static (appearance only) cues and another on dynamic (motion) cues, are compared. The motion-cue-based hand detection uses temporally accumulated edges in order to maintain the most reliable and relevant motion information. The accumulated image is fitted with ellipses in order to produce the location of the hands. The method is used to identify three hand activity classes: (1) two hands on the wheel, (2) hand on the instrument panel, (3) hand on the gear shift. The static-cue-based method extracts features in each frame in order to learn a hand presence model for each of the three regions. A second-stage classifier (linear support vector machine) produces the final activity classification. Experimental evaluation with different users and environmental variations under real-world driving shows the promise of applying the proposed systems for both postanalysis of captured driving data as well as for real-time driver assistance. © 2013 SPIE and IS&T [DOI: 10.1117/1.JEI.22.4.041119]*

## 1 Introduction

Human-centered active safety approaches recognize the importance of continuous monitoring of vehicle dynamics so that dangerous situations can be safely averted. The research in our lab for the past decade has emphasized the "looking in and looking out" approach,[1] where novel machine vision systems are introduced for observing activities of drivers/passengers in the vehicle and simultaneously capturing and analyzing the dynamic surround of the vehicle.

This paper deals with machine vision approaches for detecting hands and hand activities of a driver in video data captured in naturalistic driving conditions, reliably and robustly. "Keep hands on the wheel and eyes on the road" is a popular mantra used by instructors teaching safe driving practises. Inferring information from hand activity is especially important in the operated vehicle because it may provide vital information about the state of attentiveness of the driver. Secondary tasks in the vehicle, in particular activities involving driver's hands in the car, were shown to affect certain attention markers such as total eyes off the road.[2] Because driver distraction is a leading cause of car accidents[3], studying where the hands are and what they do in the vehicle has never been a more pressing matter.

Drivers are increasingly engaged in secondary tasks behind the wheel (23.5% of the time according to Ref. 2), which have been highly correlated with driver state and attention level. Different hand activities require different levels of visual, manual, and cognitive attention. For instance, cell-phone usage is known to significantly hinder driver awareness and reaction capabilities.[4] According to a recent survey, 37% of the drivers admit to having sent or received text messages, with 18% doing so regularly while operating a vehicle.[5] Other secondary tasks were also shown to produce increased distraction and are also prevalent. For instance, 86% of drivers report eating or drinking (57% report doing it "sometimes" or "often"), and many report common GPS interaction, surfing the Internet, watching a video, reading a map, or grooming. Knowledge of hand activity in the vehicle could result in a better understanding of driver behavior and improved assistive technology.

We first discuss existing literature on hand detection and tracking, focusing on in-vehicle hand activity recognition. These efforts are used to highlight challenges that arise in studying hand activity in the vehicle under naturalistic driving settings. A naturalistic dataset is collected and analyzed using two methods for hand detection. One method performs activity classification using motion cues and another using static (appearance) cues. Activity classification is performed in three regions of interest (ROIs): the wheel, instrument panel, and gear shift. Experimental evaluation follows, and concluding remarks with directions for future work are discussed.

## 2 Related Studies

Vision-based human hand detection is challenging, primarily because of the wide range of configurations and appearances a human hand can assume, and its tendency to occlude itself in images (self-occlusion). Low-level descriptors used in methods for the detection and tracking of hands in color or gray-scale images may fall under two broad categories, static or dynamic (motion-based) features. Each has its set of advantages and limitations; static cues, such as edges,

Paper 13234SS received Apr. 25, 2013; revised manuscript received Aug. 28, 2013; accepted for publication Sep. 6, 2013; published online Oct. 21, 2013.

0091-3286/2013/$25.00 © 2013 SPIE and IS&T

color, and texture, have gained popularity for object detection purposes and constitute one of the approaches used in this paper. We also sought a motion-based approach since it is potentially less prone to rotation, appearance variations, and occlusion cases. Nonetheless, integrating multiple cues will likely provide the optimal solution for the challenging problem at hand.

Currently, many state-of-the-art algorithms for localization of hands are designed and evaluated in lab settings. Often these methods suffer from large false positives under volatile illumination changes, cluttered background, and complex environments[6]. Recently, Mittal et al.[7] proposed multiple cue integration for hand detection from edge, skin, and geometry features, and showed state-of-the-art results on a challenging hand detection dataset. However, the hand shape detector [based on deformable part model and histogram of oriented gradients (HOG) features[8]] or tracking algorithms, such as in Ref. 9, are still limited in the cluttered and volatile in-vehicle environment.[10]

Studies relating to in-vehicle hand activity are shown in Table 1. Overall, the topic has been studied using different sensor modalities and experimental settings, as well as with different activity classes. At the basic level, user determination of hand was performed.[11,12] Leveraging thermal imagery from multiple cameras allowed for higher-level activity analysis in naturalistic settings.[13] Hand detection is commonly performed using skin-based approaches, which is highly sensitive to environmental settings[10] as well as inapplicable to naturalistic driver studies where color is not available (e.g., the Strategic Highway Research Program Naturalistic Driving Study, SHRP2-NDS). The perspective views in each of these also differ significantly, as shown in Table 2.

## 3 Hand Activity Analysis for Large-Scale Naturalistic Driving Studies

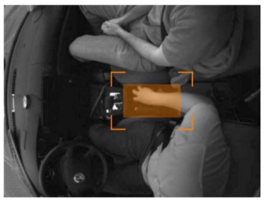
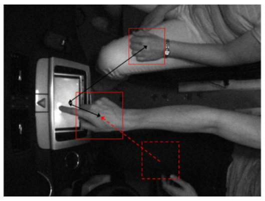
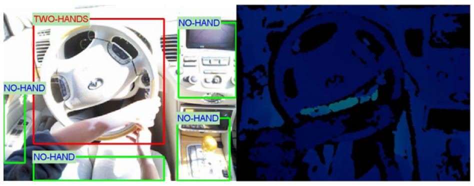
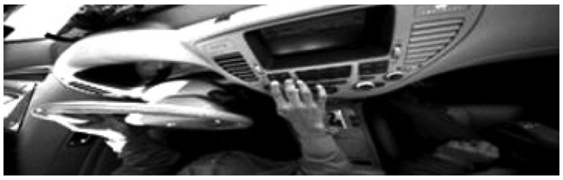
The basic "looking in and looking out" framework requires collection and analysis of a very large amount of naturalistic driving data, from a wide range of real-world driving conditions. Once such data are collected, systematic "driving ethnography" studies can be undertaken to investigate various safety-related issues.[1] A recent notable effort to study such phenomenon is the SHRP2-NDS.[16] Since fall 2010, visual data were collected for two years in six cities and 1950 vehicles. As part of a five camera unit, one camera observes the hands of the driver-wheel, gear, and instrument panel regions. As the video is recorded under very low quality and unconstrained settings, this work of hand activity recognition is directly applicable to the automated extraction of semantic gesture information from the large amount of data in the SHRP2-NDS. Automatic hand activity analysis is necessary in order to gain further insight into the process leading up to a distracted driver and the relationship between in-vehicle activities and safety. Furthermore, real-time hand detection can provide useful insight into what the driver

**Table 1** Overview of selected studies for looking at driver hand activity in a vehicle. Algorithmic approach includes hand detection method (DT) and activity classification method (CL).

| Research study | Sensor | Perspective | Activity classes | Algorithmic approach | Experimental settings |
|---|---|---|---|---|---|
| Veeraraghavan et al.[14] | One color camera | Driver side as viewed from passenger window | **Two**: driving or talking | **DT:** Skin threshold in RGB space **CL:** Bayesian eigen-image | Parked vehicle |
| Cheng et al.[13] | Two long-wavelength infrared (thermal) cameras | From over the right shoulder | **Three**: going forward, turning left, turning right | **DT:** Haar-like + Adaboost and a Kalman filter with probabilisticdata association **CL:** Hidden Markov model | Naturalistic driving |
| Tran and Trivedi[15] | Two color cameras | From over the right shoulder | **Three**: two, one, or no hand on the wheel | **DT:** Skin threshold in L*a*b space **CL:** Conditional state machine | Naturalistic driving |
| Cheng and Trivedi[11] | One color camera + LED IR | Top-down view | **Three**: driver's hand, passenger's hand, or no hand in the infotainment region | **DT:** HOG + SVM<br>**CL:** SVM | Parked and naturalistic driving |
| Herrmann et al.[12] | One color camera + LED IR | Top-down view | **Two**: driver or passenger interacting with a touch screen | **DT:** Haar-like + Adaboost and motion cues **CL:** Direction of motion | Simulator, different illumination settings are studied |
| Ohn-Bar and Trivedi[10] | One color camera and depth (Kinect) | From over the right shoulder | **Five**: hand in five regions of wheel, lap, hand rest, gear, and instrument cluster | **DT:** edge and texture features + SVM **CL:** Second-stage SVM | Naturalistic driving |
| This study | One camera, monochrome | Top-down view | **Three**: two hands on the wheel, one hand on instrument cluster, or one hand on gear shift | **DT:** edge and texture + SVM and edge-based motion features **CL:** Second-stage SVM and Euclidean distance | Naturalistic driving |

Note: HOG, histogram of oriented gradients; SVM, support vector machine.

**Table 2** Respective camera perspective views of the selected studies presented in Table 1.

| Research study | Camera perspectives |
|---|---|
| Veeraraghavan et al.[14] | |
| Cheng et al.[13] | |
| Tran and Trivedi[15] | |
| Cheng and Trivedi[11] |  |
| Herrmann et al.[12] |  |
| Ohn-Bar and Trivedi[10] |  |
| SHRP2-NDS[16] | |
| This study |  |

may intend to do, as was shown for the case of turn intent prediction.[17] It can also be used to enhance user interfaces in the car.[18]

The camera perspective of the hand view from a published sample of SHRP2-NDS is replicated in our own testbed (see Fig. 1 for a side-by-side comparison). The perspective is a top-down view of the instrument panel with a bias toward the steering wheel extending from the driver's side door to half of the passenger seat. A wide angle camera of 135 deg field of view was used. Most NDS data acquisition mechanisms crop and scale original images with multiplexing to downsize the image size for storage.

Such view may be advantageous for large data acquisition and minimal sensor cost and setup, but they pose many challenges to a vision-based analysis system. In particular, the methods must be robust to illumination changes and generalize well over users and operating modes. Additionally, hand activity states must be segmented correctly. Self-occlusion, occlusion by another hand, or by an object are common and must be addressed. In addition to the difficulty of tracking the deformable hand, the harsh visual settings usually make precise pose tracking, which is at the core of many hand activity analysis techniques, difficult.

## 4 Vision-Based Hand Activity Analysis

Two different frameworks are considered for hand activity analysis: one uses motion descriptors in a clustering framework and another uses low-level appearance descriptors in a learning framework. The former approach takes advantage of the strong motion cue that presents itself when hands are moving to track the hand. Figure 2(a) shows the potential for hand tracking using motion cues. However, motion cues can be susceptible to other movements near the ROI or illumination variation. In the latter approach, spatial constraints are introduced to address the difficult problem of hand detection and tracking. It combines detection results from individual spatial regions in order to perform activity classification. Figure 2 shows proof-of-concept of this system with three states: (1) two hands on the wheel, (2) one hand on the instrument cluster. and (3) one hand on the gear. In the following sections, we describe these approaches in more detail.

### 4.1 *Motion-Based Detection and Activity Classification*

Motion is a useful cue to leverage—as the hand transitions between regions, it produces distinctive motion cues. Additionally, motion cues can detect when there is motion within a region, such as when the hands are operating the

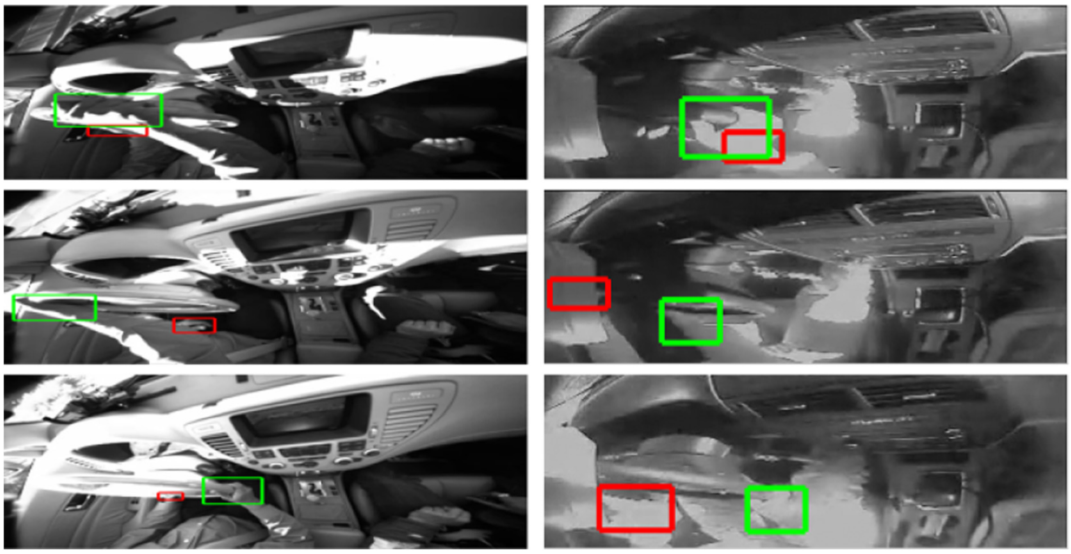


**Fig. 1** Examples of challenges for vision-based in-vehicle hand localization. Annotations: in red is the left hand and in green is the right hand of the driver. Our test bed (first three images on the left) contains a wide-angle camera that captures matching perspectives with the Strategic Highway Research Program Naturalistic Driving Study, SHRP2-NDS, hand view on the right. The view presents many challenges for automatic vision-based hand analysis techniques, such as frequent occlusion by the other hand or objects in the cabin.

wheel in the wheel region. Motion cues are more robust to appearance changes, for instance, if the perspective were to vary or when the regions differ significantly among different vehicles and users. Nonetheless, motion algorithms are usually sensitive to illumination changes.

Our approach incorporates edge features for tracking driver's hands. The flow diagram of the algorithm is illustrated in Fig. 3. A tight ROI was essential to prevent false motion cues around the cabin. To track the hand, an edge image $E_t$ is first extracted using the Canny edge detector for each input image frame. Then, the hand motion is detected and accumulated using edge image differencing over the extent of the continued motion. At each frame $t$, the motion image $M_t$ and the accumulated motion image $AM_t$ were computed as

$$M_t = E_t - \sum_{i=t-a}^{t-1} E_i,$$

$$AM_t = \cup_{i=t-b}^{t} M_i,$$

where $1 < a \leq \{w\text{:windowsize}\}$ and $b \geq a$. The window size represents how far back in time to look in order to recognize new motions and ignore older motions. A pixel that was previously stationary but now moved results in a positive value in $M_t$, which is thresholded to a binary image. Next, the OR operation is applied over the window of time (we use $a = 10$, $b = 40$).

The resulting accumulated motion image is then segmented into connected motion regions, and a threshold on those regions is used to remove some of the motion artifacts due to illumination changes. Finally, ellipses are fit over filtered prominent motion regions. The ellipse parameters represent the characteristics of segmented motion history and the centroid determines the proposed position of the hand (shown in blue in Fig. 3). Based on the hand tracking

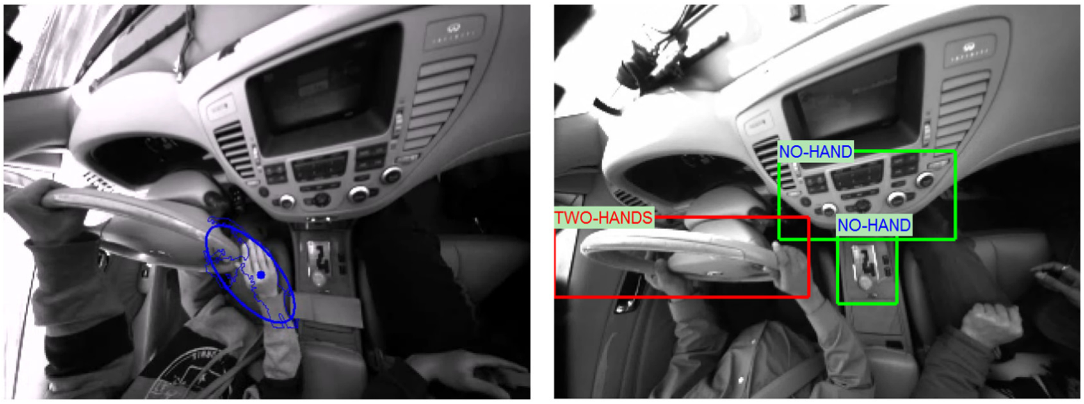


**Fig. 2** Proposed framework for analyzing hand activities through (a) motion-cues from accumulated edge differencing and (b) a static/appearance only approach where a hand model is learned for three regions of interest in the vehicle (wheel, instrument cluster, and gear shift) and integrated using a second-stage classifier.

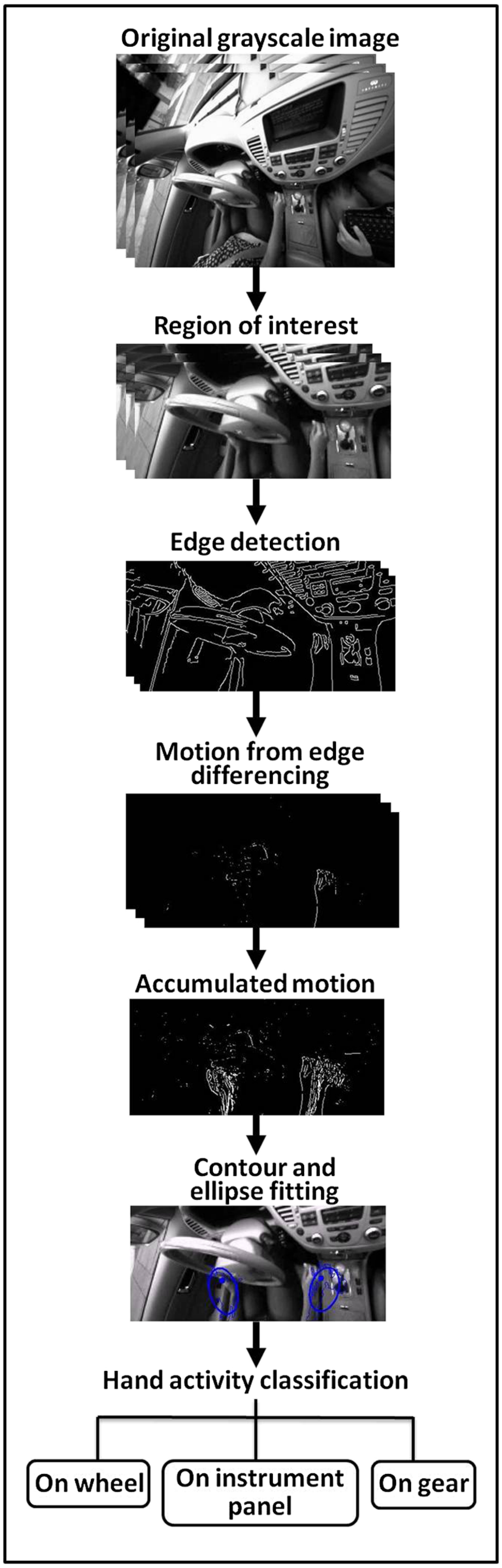


**Fig. 3** Schematic diagram of hand activity detection based on motion cue analysis. Edge images were generated from input images using Canny edge detector and prominent motion regions are extracted by thresholding. Segmented overlapping motion of current and previous frames indicate the hand position (shown as a circle inside the ellipse).

output, driver hand activity is classified into three classes of interest—on wheel, gear, or instrument cluster—using the minimum Euclidean distance from the tracked rightmost hand to the defined center location of each region.

### 4.2 *Appearance-Based Detection and Activity Classification*

Due to harsh visual setting, a sliding window-based detector trained on hand instances was shown to be prone to false positive detection rates on our data. Instead, we can leverage an assumption that the hands can only be found in a small set of regions (three in this case) and integrate cues from these predefined ROIs in order to gain robustness.[10] The general flow of this scheme as it applies to our specific three activity states case study can be seen in Fig. 4. This scheme prunes false positives by using the confidence scores from individually learned models for hands in each region in order to produce the final activity classification.

We investigated several features for the feature extraction process and found them to vary in performance within regions. We compared the following features:

HOG: Based on Ref. 19. Since the same regions vary in size among the subjects due to perspective changes that occurred over time, we resize wheel images to size $140 \times 240$, instrument cluster images to $90 \times 190$, and gear shift images to $90 \times 60$. Cell size used was 8 and number of orientations was 9.

Modifed HOG: A relatively low dimensional descriptor explored in Ref. 10. This descriptor benefits from different resolutions of cells. The parameters specify the number of cells over the image, and so we use one cell and eight cells with eight histogram bins to produce the $8 + 512 = 520$ dimensional feature set.

Dense scale-invariant feature transform (SIFT)[20] + principle component analysis (PCA): Computed with a spatial binning size of $3 \times 3$ pixels, and PCA is used to reduce the final descriptor size for each pixel to a 20-dimensional vector.

Furthermore, Global features are the mean, median, and variance of the pixel intensities, and Difference of HOG features is computed by subtracting left and right modified HOG features in an image region, as described in Ref. 10. The GIST descriptor[21] is also compared to the aforementioned.

A linear support vector machine (SVM) classifier is trained for binary classification of hand presence/no-presence in each region. A second-stage classifier provides the final classification into the specified states (e.g., two hands on wheel, one hand on gear, etc.). This classifer leverages the fact that there are only a discrete number of possible states (three in this case) defined for the hand, allowing for higher-level reasoning of the activity in a current frame. Second, due to the unbalanced nature of the occurrences of activity classes (see Sec. 5.1), we use a biased penalties SVM (Ref. 22) and the LIBSVM (Ref. 23) implementation. This framework can correctly classify hand presence in smaller regions, such as the gear, with good reliability, and the larger, more difficult areas are more prone to erroneous detections. In this lies the advantage of integrating multiple small ROIs to infer the correct hand location. Incorporating temporal information was shown

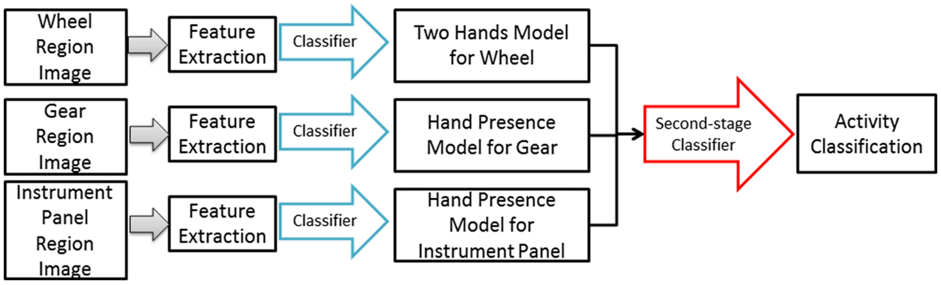


**Fig. 4** Schematic diagram of hand activity detection based on appearance cues analysis. As opposed to training a model for hand shape or appearance and running a sliding window detector, we train a binary classifier for each region. The model is trained to distinguish a foreground (the hand) in a region of interest. Because the hands appear differently in each of the regions and each region is prone to different visual challenges, it is beneficial to learn a unique model for each region. These output probability scores, which are integrated using a second-stage classifier (linear SVM), in order to produce a high-level representation of the scene in terms of the final activity classification.

to significantly improve the results in transition times in Ref. 11.

## 5 Experimental Evaluation

### 5.1 *Dataset*

The two techniques proposed are evaluated on nine video sequences. Each of the collected nine video sequences contains different drivers and can be described with the following attributes: weather conditions, illumination effects, and background clutter (Table 3). Weather condition is the main indication for the illumination volatility in the scene. Sunny conditions resulted in more difficult settings, with shadows of inside and outside objects producing false positive detection of hand activity. Background clutter was introduced in two videos, where different objects (cables, cellphones, wallets, etc.) were placed and included in the gear shift region (see Fig. 5, second row). Ground truth for evaluation of hand activity is available from manual annotation of the presence of hands in the three regions, producing a total of 16,209 instances. Table 3 shows the instant count of the hand in each of the regions. As in naturalistic driving settings, the classes are unbalanced. In our case, the most occurring class is the "two hands on the wheel" activity. Furthermore, the dataset is challenging since the wheel region is the most difficult out of the three for activity detection.

### 5.2 *Evaluation Metric*

For performance analysis, we use the normalized accuracy measure

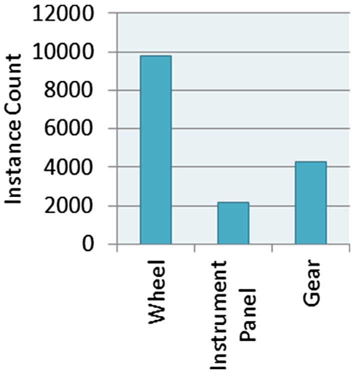
$$CCR = \frac{1}{K} \sum_{c=1:K} p_c,$$

where $K$ is the total number of classes, and $p_c$ denotes the percentage of correctly matched instances for class c. This normalizes for the unbalanced class instances within the three regions.

### 5.3 *Discussion*

The motion-based algorithm provides activity classification for each of the nine videos. The results are summed in Table 4. The algorithm runs at ~80 frames per second. Per-frame classification is low (at 47.3% correct classification accuracy) because the ellipses are mostly accurate in hand transitions. But in the cases where there is little motion information (when a hand is interacting with the instrument cluster or gear shift), there will usually be motion coming

**Table 3** Dataset statistics for activity instances for each class.

| Subject | Weather | Illumination effects | Background clutter | Total instances |
|---|---|---|---|---|
| 1 | Sunny | Large | No | <br> |
| 2 | Sunny | Large | No | |
| 3 | Overcast | Small | No | |
| 4 | Sunny | Large | Yes | |
| 5 | Sunny | Large | No | |
| 6 | Sunny | Large | Yes | |
| 7 | Overcast | Small | No | |
| 8 | Overcast | Small | No | |
| 9 | Overcast | Small | No | |

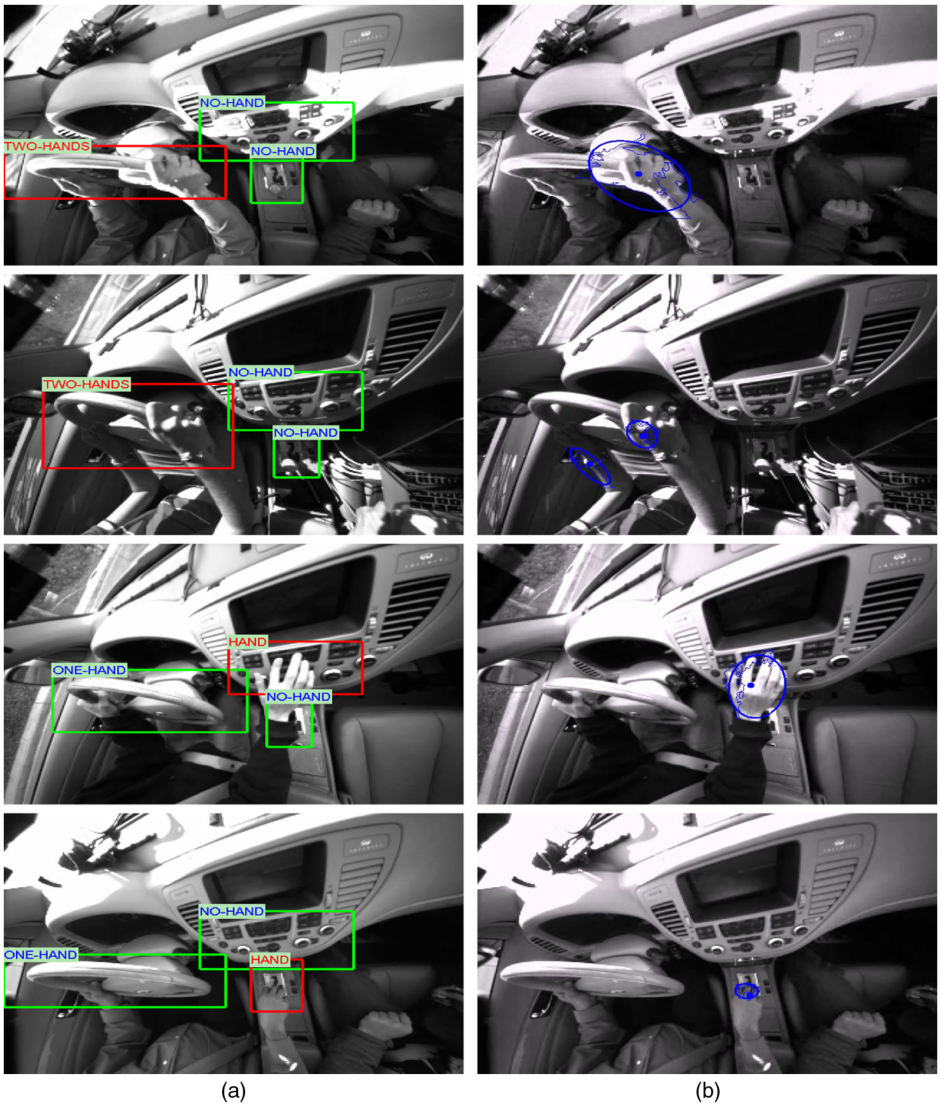


**Fig. 5** Correct classification results using the static (appearance) cues approach (a) and dynamic (motion) cues approach (b).

**Table 4** Confusion matrices for (a) appearance-based algorithm, normalized accuracy 74.3 and (b) motion-based algorithm, normalized accuracy 47.3.

| **Class** | **Predicted** | | |
|---|---|---|---|
| | **Wheel** | **IC** | **Gear** |
| (a) Static/appearance algorithm | | | |
| Wheel | **0.76** | 0.24 | 0 |
| Instrument cluster (IC) | 0.12 | **0.83** | 0.05 |
| Gear | 0.23 | 0.13 | **0.64** |
| (b) Dynamic/motion algorithm | | | |
| Wheel | **0.91** | 0.04 | 0.05 |
| IC | 0.47 | **0.34** | 0.19 |
| Gear | 0.82 | 0.01 | **0.17** |

from the other hand or the wheel region, or from illumination artifacts (shown in Fig. 6). Therefore, a temporal model should be used. Preliminary results show a significant increase in accuracy by incorporating location and motion statistics over a window of frames.

For the appearance-based classification scheme, we first evaluate each descriptor and best performing descriptor combinations in each region individually, where the task is to simply detect hand presence (or two hands on the wheel in the wheel region). Results for this two-class problem are shown in Fig. 7 using a leave-one-subject-out cross-validation. The legend for the descriptors is shown in Table 5. Mean normalized accuracy is plotted over the nine tests for each feature, as well as the standard deviation. Top performing descriptors will be used to produce the final activity classification in the three regions. We notice the lower performance within the large and difficult wheel region, where using a modified HOG + Global + Difference of HOG produces good results in a relatively fast feature extraction process. GIST takes the longest to compute, although it was shown to produce good results in the small regions. The final activity classification results are shown in Table 4, produced

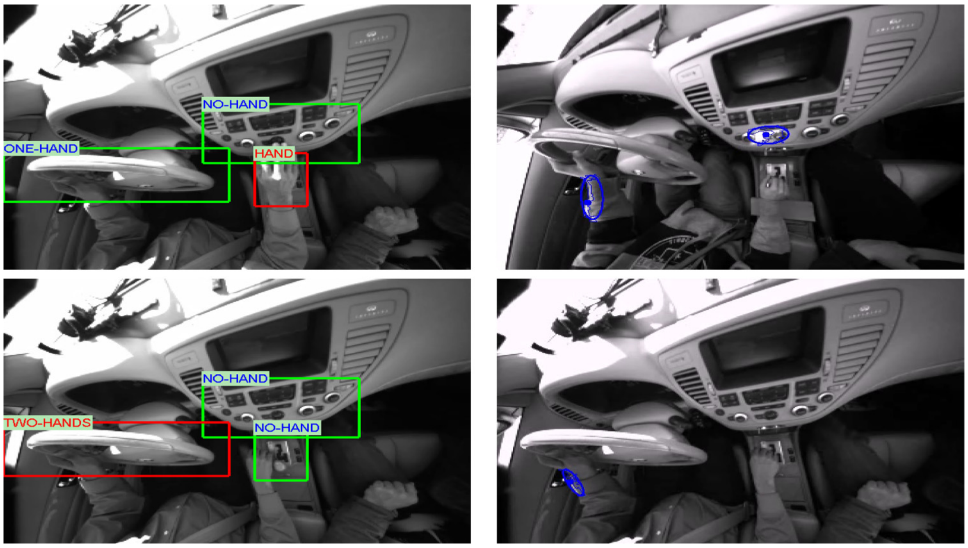


**Fig. 6** Incorrect classification results using the two techniques. Transitions between regions produce incorrect classification for the appearance-based scheme. Both methods still exhibit sensitivity to illumination changes, but the motion-based method is significantly more sensitive. Furthermore, the method is biased toward the wheel region as there may not be distinctive motion-cues when the hand is in the gear or instrument cluster regions.

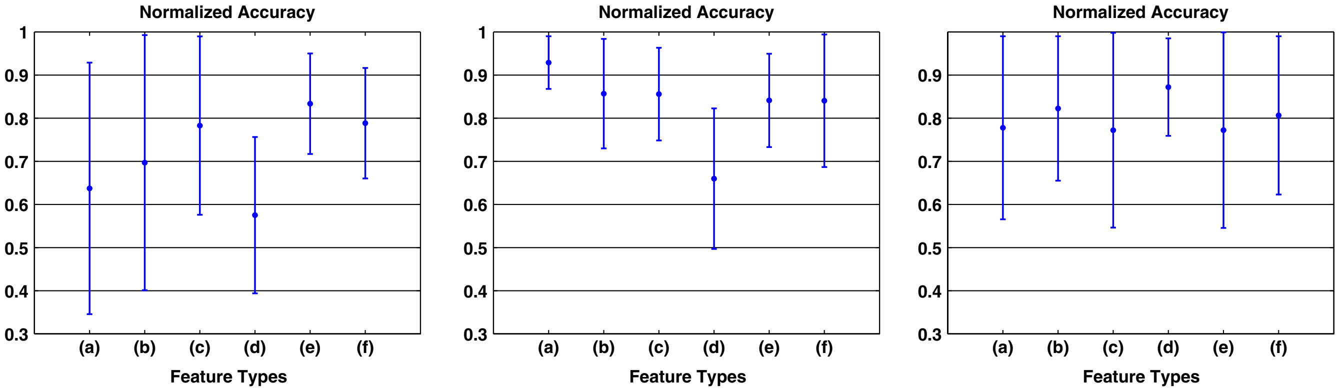


**Fig. 7** Performance results using the appearance algorithm with leave-one-subject-out cross-validation. Each region is tested in a two-class problem of hand or no hand for the instrument cluster or gear shift regions, and two hands or not for the wheel region. Mean accuracy is plotted over the nine tests for top performing features, as well as the standard deviation. Feature types used are labeled according to the legend in Table 5. Notice how different descriptors perform differently among the regions.

**Table 5** Top performing descriptors and their labels in Fig. 7.

| Feature label | Detail |
|---|---|
| (a) | GIST |
| (b) | Histogram of oriented gradients (HOG) |
| (c) | Modified HOG |
| (d) | Dense SIFT + PCA |
| (e) | Feature (c) + Global + Difference of HOG |
| (f) | Features (e) + (d) |

using a twofold cross-validation where half the subjects were used for training and the other half for testing, and then reversed. The confusion matrix was generated by averaging the results of these two tests. Although certain descriptor selection in the appearance-based scheme underperforms the results given by the motion-based scheme, choosing the right descriptors leads to a 74.3% normalized accuracy. This scheme mainly benefits from the second-stage classifier. Integrating the static-appearance and dynamic-motion schemes is an important future work.

## 6 Concluding Remarks

Robust, vision-based systems for studying driver behavior have many important applications. Such algorithms could facilitate the discovery of new insights on driver behavior using automatic analysis of large amounts of data. Furthermore, they can be used to design novel and safe advanced driver assistance systems.

In this paper, we presented an analysis of existing literature relevant to driver hand activity classification in terms of sensor, perspective, descriptors, activity classification methodology, and the extent of the experimental evaluation. Furthermore, we highlighted the challenges such a task poses, especially for vision-based system relying on monocular, monochromatic input. As a case study, we studied two algorithms, each leveraging different cues, on a challenging dataset with a difficult perspective and varying environmental settings.

Although certain appearance-cues were shown to perform better on our dataset than the motion-cues, reliable motion-cues have the potential to provide certain advantages, for instance, more robustness to scene and perspective changes. Therefore, further study of extraction methods is needed. Temporal models can provide a probabilistic model for the state of the hands, building on top of the low-level motion-cues and leading to a more robust classification scheme. Additionally, preliminary analysis using high-quality optical flow,[24] although computationally expensive, showed more robustness to illumination changes. Future work should also include appropriate integration schemes of static and dynamic cues, as these are expected to be complementary.

Further testing under challenging settings that were not pronounced in our dataset, such as heavier occlusion by objects in the scene, lower-resolution images, and larger variations in the viewing angle, and its effects on the performance of the two algorithms needs to be performed. We are already planning extensive studies at two different locations in the United States with different vehicles and drivers to pursue such robustness evaluation. Studying the effects of extending the vocabulary of activities for more than the three proposed in this paper, including the study of hand-object interaction and temporal hand gestures relating to driver intent, will be useful for better studying driver behavior. Furthermore, such systems can be incorporated with other machine vision systems for looking inside the vehicle, such as head pose systems[25] in order to provide a more comprehensive monitoring of driver activity. Advanced driver assistance systems can integrate cues coming from inside the vehicle and cues from the dynamic surround of the vehicle to produce improved recommendation and assistance.

## Acknowledgments

We acknowledge support of the UC Discovery Program and associated industry partners. We would like to thank Toyota Collaborative Safety Research Center for its sponsorship. We are especially grateful to Dr. Pujitha Gunaratne of Toyota for his support and many useful inputs. We also thank our UCSD LISA colleagues, in particular Dr. Cuong Tran and Mr. Minh Van Ly, and are also thankful to Mr. Greg Dawe for his contribution on the design and installation of the video module. Finally, the authors would like to thank the reviewers and editor for their constructive comments for improving the quality of this paper.

## References

1. M. M. Trivedi, T. Gandhi, and J. McCall, "Looking-in and looking-out of a vehicle: computer-vision-based enhanced vehicle safety," *IEEE Trans. Intell. Transp. Syst.* **8**(1), 108–120 (2007).
2. S. Klauer et al., "An analysis of driver inattention using a case-crossover approach on 100-car data: final report," Technical Report DOT HS 811 334, National Highway Traffic Safety Administration, Washington, D.C. (2010).
3. J. Tison, N. Chaudhary, and L. Cosgrove, "National phone survey on distracted driving attitudes and behaviors," Technical Report DOT HS 811 555, National Highway Traffic Safety Administration, Washington, D.C. (2011).
4. R. L. Olson et al., "Driver distraction in commercial vehicle operations," Technical Report FMCSA-RRR-09-042, Virginia Tech Transportation Institute, Washington, D.C. (2009).
5. T. H. Poll, "Most U.S. drivers engage in 'distracting' behaviors: poll," Technical Report FMCSA-RRR-09-042, Insurance Institute for Highway Safety, Arlington, Virginia (2011).
6. E. Ohn-Bar and M. M. Trivedi, "The power is in your hands: 3D analysis of hand gestures in naturalistic video," in *Comput. Vis. and Pattern Recognit Workshops-AMFG* (2013).
7. A. Mittal, A. Zisserman, and P. Torr, "Hand detection using multiple proposals," in *British Machine Vision Conf.* (2011).
8. P. F. Felzenszwalb et al., "Object detection with discriminatively trained part-based models," *IEEE Trans. Pattern Anal. Mach. Intell.* **32**(9), 1627–1645 (2010).
9. L. Zhang and L. V. D. Maaten, "Structure preserving object tracking," in *IEEE Conf. Computer Vision and Pattern Recognition* (2013).
10. E. Ohn-Bar and M. M. Trivedi, "In-vehicle hand localization using integration of regions," in *IEEE Intelligent Vehicles Symp.* (2013).
11. S. Y. Cheng and M. M. Trivedi, "Vision-based infotainment user determination by hand recognition for driver assistance," *IEEE Trans. Intell. Transp. Syst.* **11**(3), 759–764 (2010).
12. E. Herrmann et al., "Driver/passenger discrimination for the interaction with the dual-view touch screen integrated to the automobile centre console."*Proc. SPIE* **8295**, 82950W (2012).
13. S. Y. Cheng, S. Park, and M. M. Trivedi, "Multi-spectral and multi-perspective video arrays for driver body tracking and activity analysis," *Comput. Vis. Image Underst.* **106**(2), 245–257 (2007).
14. H. Veeraraghavan et al., "Driver activity monitoring through supervised and unsupervised learning," in *IEEE Conf. Intelligent Transportation Systems* (2005).
15. C. Tran and M. M. Trivedi, "Driver assistance for keeping hands on the wheel and eyes on the road," in *IEEE Conf. Vehicular Electronics and Safety* (2009).
16. K. L. Campbell"The SHRP2 naturalistic driving study: addressing driver performance and behavior in traffic safety," TR News 282, 2012, http://www.shrp2nds.us/.
17. S. Y. Cheng and M. M. Trivedi, "Turn-intent analysis using body pose for intelligent driver assistance," *IEEE Trans. Pervasive Comput.* **5**(4), 28–37 (2006).
18. E. Ohn-Bar, C. Tran, and M. M. Trivedi, "Hand gesture-based visual user interface for infotainment," in *ACM Automotive User Interfaces and Interactive Vehicular Applications* (2012).
19. N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE Conf. Computer Vision and Pattern Recognition* (2005).
20. D. G. Lowe, "Object recognition from local scale-invariant features," in *IEEE Intl. Conf. Computer Vision* (1999).
21. A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *Int. J. Comput. Vis.* **42**(3), 145–175 (2011).
22. F. R. Bach, D. Heckerman, and E. Horvitz, "Considering cost asymmetry in learning classifiers," *The J. Mach. Learn. Res.* **7**, 1713–1741 (2006).
23. C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM Trans. Intell. Syst. Technol.* **2**(3), 27:1–27:27 (2011).
24. D. Sun, S. Roth, and M. Black, "Secrets of optical flow estimation and their principles," in *IEEE Conf. Computer Vision and Pattern Recognition* (2010).
25. S. Martin, A. Tawari, and M. M. Trivedi, "Monitoring head dynamics for driver assistance: a multiple perspective approach," in *IEEE Conf. Intelligent Transportation Systems* (2013).

**Eshed Ohn-Bar** received his MS degree in electrical engineering from the University of California, San Diego (UCSD), in 2013. He is currently working toward a PhD degree in electrical engineering with specialization in signal and image processing at the Computer Vision and Robotics Research Laboratory and Laboratory for Intelligent and Safe Automobiles (LISA) at the UCSD. His research interests are in machine learning and computer vision.

**Sujitha Martin** received her BS degree in electrical engineering from the California Institute of Technology in 2010 and her MS degree in electrical and computer engineering from the UCSD in 2012. She is currently pursuing her PhD degree at UCSD LISA. Her research interests are in computer vision, machine learning, human-computer interactivity, and gesture analysis. She is a recipient of UCSD ECE Department Fellowship during 2010 to 2011. Her poster presentation, titled "Optical flow based head movement and gesture Analyzer (OHMeGA)," received an honorable mention at the 32nd Annual Research Expo 2013 held by UCSD Jacobs School of Engineering.

**Mohan Manubhai Trivedi** is a professor of electrical and computer engineering and the founding director of the computer vision and robotics research laboratory and laboratory for intelligent and safe automobiles at the University of California, San Diego. He and his team are currently pursuing research in machine and human perception, machine learning, human-centered multimodal interfaces, intelligent transportation, driver assistance and active safety systems. He serves as a consultant to industry and government agencies in the U.S. and abroad, including the National Academies, major auto manufacturers and research initiatives in Asia and Europe. He is a fellow of IEEE (for contributions to intelligent transportation systems field), a fellow of the International Association of Pattern Recognition (IAPR) (for contributions to vision systems for situational awareness and human-centered vehicle safety), and a fellow of SPIE (for distinguished contributions to the field of optical engineering).