Manuscript accepted for publication in Journal of Transportation Safety & Security, 2017.

A 3D Body Posture Analysis Framework During Merging And Lane Changing Maneuvers

A. Kondyli[a], A. Barmpoutis[b], V. P. Sisiopiku[c],

L. Zhang[d], L. Zhao[b], M. M. Islam[c], S. S. Patil[a], and S. Rostami Hosuri[c].

Abstract

Although significant advances have been done with respect to vehicle technology and roadway construction, driver behavior remains the number one contributing factor of traffic crashes worldwide. Studies show that one of the major causes of crashes is driver inattention, which may occur when drivers are involved with secondary activities (e.g., texting, talking on the phone, or eating), and when they fail to follow the cues of the surrounding environment while driving. The objective of this study was to develop a method that monitors driver body posture and movements inside the cabin and test it among different drivers when performing merging and lane changing maneuvers, since these types of maneuvers require significant body movement and may also result in unsafe situations. The developed method was applied in a naturalistic setting where 35 drivers were invited to participate. Participants' 3D body posture was recorded with the use of a low-cost infrared depth sensor (Microsoft Kinect). Participants' eye gaze was also recorded with the help of an eye-tracking equipment. This paper presents analysis results of 3D body posture in conjunction with the eye tracking information during 236 merging and 287 lane changing maneuvers.

[a] Department of Civil, Environmental and Architectural Engineering, University of Kansas, 1530 W. 15th St., 2159A Learned Hall, Lawrence, KS 66045
[b] Digital Worlds Institute, University of Florida, PO BOX 115800, Gainesville, FL 32611-5800
[c] Department of Civil, Construction, and Environmental Engineering, 1075 13th St S, Hoehn 311, University of Alabama at Birmingham, Birmingham, AL 35294-4440
[d] Department of Civil and Coastal Engineering, University of Florida, 365 Weil Hall, Gainesville, FL 32611-6580

Keywords

Driver Alertness; Driver Behavior; Merging; Lane Changing; Body Posture; Eye Tracking.

1. Introduction

Despite the advances in vehicle manufacturing technology and roadway construction and design, a large proportion of traffic crashes are still due to driver error (*World Health Organization-WHO, 2004*). According to the World Health Organization (WHO), annually there are over 1.2 million traffic fatalities and over 20 million serious injuries worldwide. These traffic fatalities account for 23 percent of all injury deaths worldwide. WHO (2004) also states that approximately 90 percent of the road crashes are related to human error. In the US, the 100-car naturalistic study sponsored by the National Highway Traffic Safety Administration (NHTSA) analyzed approximately 2,000,000 vehicle miles driven by 241 drivers for 12 to 13 months, and observed a total of 69 crashes and 761 near-crashes. The authors concluded that driver inattention is the cause of about 80 percent of crashes and 65 percent of near crashes (Dingus et al. 2006). These percentages are significantly higher than previous literature results, where driver distraction was found to be a contributing factor to 25 percent of crashes (Hendricks et al., 2001; Campbell et al., 2003).

Given that driver error constitutes the primary reason for the majority of traffic crashes, elimination of this factor through the adoption of autonomous and connected vehicle technology is expected to improve traffic safety. The USDOT's connected-vehicle research program uses technologies such as advanced wireless communications, on-board computer processing, advanced vehicle-sensors, GPS navigation, and smart infrastructure, to identify imminent road hazards and warn the drivers accordingly (USDOT, 2011). A number of crash avoidance systems have been established to date, such as emergency stop warning, forward collision warning, intersection movement assistance, blind spot and lane change warning, and do not pass warning. Additional advanced (or intelligent) driver assistance systems (ADAS) designed to provide added traffic safety are already in place (Shaout et al. 2011). These systems do not involve inter-vehicle communication, and are designed to provide assistance or warning to drivers by considering the longitudinal position of the vehicle or other vehicle-related components. Examples of ADAS applications include automatic parking, adaptive light control, night vision, lane change assistance, traffic sign recognition, collision avoidance system, lane departure warning system, and hill descent control. Apart from these systems that focus on the vehicle, there are limited systems already in place that are designed to monitor the driver's eye movement for inattention or drowsiness using LED sensors.

In vision-based systems that involve understanding of driver intentions and actions (e.g., inattention or distraction states) research studies focus primarily on the head and face of the driver. For instance, Tijerina et al. (2005), Trivedi et al. (2007) and McCall et al. (2007) analyzed head pose and gaze for identifying and predicting driver's intent to change lanes. Tijerina et al. (2005) observed the eye glance of various drivers using a face camera, while executing lane changes with an instrumented passenger vehicle and an instrumented van. The authors produced link diagrams showing the probabilities of a glance to a specific location (e.g., right/left mirror, road ahead, center mirror, etc.) 10 seconds prior to the lane change event. Tijerina et al. (2005) concluded that drivers did not always check their mirrors or turned their heads during the 10 seconds before starting the lane change.

Research has also studied the hand position and grasp in conjunction with head monitoring for lane change intent analysis and prediction (Cheng and Trivedi, 2006) or for driver distraction monitoring (Tran and Trivedi, 2009). Tran and Travedi (2010) presented a system for tracking the 3D body movement combined with head pose tracking. The authors tested their system in a simulation environment and obtained preliminary results related to body posture and lane changing activity. Although the experimental platform appears promising, their results to date are limited and do not consider potential differences between various driver groups. Recently, Kondyli et al. (2013) developed a 3D framework for exploring drivers' body activity using depth sensors and used a small set of data from a pilot study to show that the proposed approach captures significant differences between drivers' body movements while performing merging and lane changing maneuvers.

In summary, the literature review reveals that a significant amount of research has been involved with the development of advanced driver-assistance systems; however, most of these systems rely on the automobile position and do not necessarily consider the drivers actions. Apart from that, the lane trajectory and position of the vehicle could potentially differ from the driver's intent to change lanes. In addition safety research has focused on eye tracking as a means of capturing driver's attention, fatigue, or drowsiness; however, the entire body posture of drivers when performing a maneuver as well as differences in postures between various groups of drivers may also reveal behaviors that contribute to unsafe driving conditions.

The objective of this study is to develop a framework for modeling driver body posture and actual body movements, and in conjunction with the eye fixation of the drivers, test this framework under different traffic conditions. For the purposes of this research, the investigation of the driver body posture and eye fixation focused primarily on merges and lane changes, since these types of maneuvers are typically associated with intense body motion and eyes-off-the-road instances. The findings can assist in establishing monitoring guidelines for advanced driver assistance systems that take into account the driver's body position and movements, rather than considering solely the vehicle position relative to other vehicles on the road. The results can also assist in developing appropriate alert mechanisms for increasing driver alertness.

2. Methodology

2.1 Data Collection and Processing

The data used in this study were collected using an instrumented vehicle with cameras (front and rear), GPS, accelerometer, depth camera (in cabin), and eye tracker (head-mounted). The instrumented vehicle used was a Honda Pilot SUV, owned by the University of Florida-Transportation Research Center (TRC). The vehicle has a Honeywell Mobil Digital Recorder (HTDR400) system. This system has two digital cameras, which capture front and rear view video. The video is stored in the hard drive of the HTDR400 system. The instrumented vehicle is also equipped with a GPS. For the depth data collection the research team used the infrared structure-light sensor embedded in Kinect™ by Microsoft. The resolution of the depth sensor was 320×240 pixels at 30 frames per second and was calibrated so that it recorded depth in the range from 0.4m to 3.0m, which was adequate to capture the motion of the driver with the sensor mounted on the upper right side of the cabin. The Mobile Eye-XG eye-tracking equipment by ASL was used to capture drivers' gaze, using two high-resolution digital cameras placed on the lightweight head-mounted pair of glasses.

A total of 35 drivers were selected to participate in the field data collection based on an eligibility-screening questionnaire. Eligible drivers (i.e., with valid drivers' license and car insurance) were chosen based on their age and gender, in order to have a broad selection of participants. Each of these drivers completed a pre-driving questionnaire before each session. The pre-driving questionnaire contained several multiple-choice questions related to their driving habits, such as driving frequency to work/school, total duration of driving on weekdays, desired speeds on urban streets, lane changing frequency, and frequency of involvement in secondary tasks while driving (texting, eating, etc.).

The field data collection was conducted between February and March of 2014 and participants drove during morning (AM), midday, and afternoon (PM) peaks. Each participant drove for approximately 2 hours along a pre-selected route in Gainesville, FL (Figure 1). The route consisted of an 8 km (5.0 mi) section along I-75 in the southbound (SB) and northbound (NB) directions, and a 1.6 km (1 mi) long arterial segment (Archer Road, EB and WB directions). The freeway segment along I-75 has three lanes per direction and a posted speed limit of approximately 110 km/h (70 mi/h). The arterial segment has three through lanes per direction, several median openings, and four signalized intersections. During the two-hour experiment, participants were asked to drive along the preselected route shown in Figure 1 several times; hence multiple data samples were collected per driver. To minimize any bias in collecting natural driving behavior observations, participants were not aware of the objective of this research. A researcher accompanied the drivers at all times while performing the experiment. The researcher was sitting at the back seat and had minimum interaction with the participant, primarily to guide him/her along the driving route. No further interaction or instructions (e.g., change lanes) were given to the participants.

Figure 1 goes here

The researchers analyzed all files obtained through the field experiment. The in-vehicle video provided useful information on the subject vehicle environment while driving. This video was available for most of the participants. The eye-tracking glasses that captured drivers' gaze were not available for all participants, because some did not feel comfortable wearing the glasses while driving. In addition, some of the eye-tracking videos were not used due to inability to calibrate the files. As a result, eye-tracking video data from 13 participants were eventually used. Kinect sensor video files from 27 participants were considered in the analysis, as it was not possible to process the remaining files due to technical problems. The data sequences obtained from the Kinect sensor are stored in the online Driver's Motion Depth DataBase (DMDDB), which is available at: http://research.dwi.ufl.edu/dmddb.

It should be noted that processing of all video files collected was a very tedious task that required significant time and effort. An issue that was encountered during the data reduction phase involved the synchronization of the various types of video data, in order to check all in-vehicle/driver-related and vehicle environment-related information. The Kinect videos and the eye-tracking videos started approximately at the same time, so these were more or less aligned. However, the in-vehicle videos would start as soon as the car ignition was on, so a significant effort was made to synchronize these videos with the rest.

2.2 Data Analysis

Two data analysis approaches were followed as part of this study. The first one was a quantitative approach that was based on: (a) obtaining time-series of body posture during driving, (b) deriving an "average" driver behavior, and (c) measuring deviations from that average for all study participants. The second approach looked into a qualitative evaluation of driving conditions and how driver posture is related to those. Both approaches considered only the merging and lane changing maneuvers that took place during the experiments on the freeway and the arterial segment. This was done primarily because these types of maneuvers are associated with intense body motion and increased safety risk as drivers interact with others while negotiating a move to their target lane. Other driving situations, such as car following were beyond the scope of the study and were not analyzed.

For the purposes of this study, video clips of the merging and lane changing maneuvers were isolated from the entire clips as part of the data reduction process. Merging maneuvers occurred at three ramp junctions of I-75 with Archer Rd (NB), Newberry Rd (SB) and Williston Rd (NB) (Figure 1). The start and ending points of the merging maneuvers vary by junction but for each junction are fixed; i.e., the maneuver starts when the vehicle enters the acceleration lane and ends when the vehicle crosses a specific point of the freeway, just after the end of the acceleration lane. This was done to facilitate comparison of the maneuvers between the different drivers. The duration of the merging clips is approximately 30 seconds (on average as it depends on the vehicle speed). For the lane changes a different approach was undertaken since these maneuvers are discretionary. The calculation of the entire lane changing maneuver duration requires going back before the actual lane changing occurs, and identify when drivers have actually thought and decided to perform the maneuver. This thought process was difficult to track, but it was done by checking whether the driver actually looked through the mirror in the eye-tracking video.

2.2.1 Quantitative Analysis

For the quantitative approach, the depth sequences for each merging and lane changing maneuver were used to study the body posture for the entire duration of these maneuvers. An algorithm was developed that located and identified the regions of the arms of the drivers using an image segmentation technique applied to the depth frames of the dataset (Figure 2).

Figure 2 goes here

The area of machine learning has rich literature that presents and evaluates supervised learning algorithms that employ different types of classifiers (Haykin, 2009). Artificial neural networks and support vector machines are just a few examples that use classifiers with simple polynomial forms, which lead to efficient algorithms with linear complexity. The classifier is typically defined as a parametric function into two classes based on the sign of the classifier's value. The unknown parameters of the classifier can be estimated by a supervised training process that fits a set of training samples and their corresponding desired responses to the classifier using an optimization process. For the purposes of our experiments a multi-layer feed-forward neural network was employed. The implemented neural network had 3 layers with 10 neurons in the input and intermediate layer, and two neurons in the output layer. The network was trained by performing least squares fitting to a manually defined training set using the Levenberg-Marquardt optimization method (Lawson, 1974) and the back-propagation algorithm (Haykin, 2009).

The effectiveness of the learning algorithm, and as a consequence the performance of the classifier, depends on the type of features used as the input vector and their descriptive capabilities, in terms of how rich is the information that they contain. Each depth frame is given as a 2D matrix, in which each element $d(i, j)$ represents the depth value of a 3D point whose perspective projection falls into the (i, j) pixel location at the frame lattice. These raw features, i.e. d, i, and j are functions of the original point in the physical 3D space of the vehicle's cabin. More specifically, the elements i and j are the projections of the x and y coordinates of the original 3D point and their relationship is defined by the equation of perspective projection. In addition to the reconstructed 3D coordinates of each pixel, the corresponding normal vectors were also calculated. Each of the depicted points lie on the surface of an object, which is described by various structural characteristics. The normal vector at the particular 3D point on the surface of the depicted object describes the local orientation of the surface and was calculated using a discrete approximation, which is expressed as a function of the depth difference Δd from the neighboring pixels. The aforementioned process yielded a 5-dimensional feature vector set that consists of *x,y, d(i, j), Nx, Ny* for every pixel of each the depth frame in the dataset, which was used as input to the classification method.

The calculated 3D point coordinates and normal vectors were with respect to the coordinate system of the depth camera. However, the depth sensor could have been installed in each data collection session at a slightly different location and orientation in the vehicle's cabin. As a result, the feature vectors from two different depth sequences may correspond to two different coordinate systems and therefore are not directly comparable. To address this issue, the research team extended further the feature extraction method by introducing a camera calibration step. In the first frame of each depth sequence, the location and orientation of the driver's seat was estimated using simple template matching with the depth image of the driver's seat, which was manually cropped from the dataset.

It should be noted that the main reason for employing a classification process in this paper was to assist us in the annotation/labeling of the recorded data by expediting the manual process. Since the experimental setup was fixed, i.e. same vehicle, same camera, almost fixed perspective, the complexity of the problem is significantly reduced. For this reason the training of the classifier is performed with 100% of the manually annotated data, which serve as key frames in the recorded depth image sequences, and the test frames are the remaining frames between the keyframes. Therefore, the supervised classification process is reduced to a key-frame based interpolation process.The reader is referred to Barmpoutis et al. (2015) for a more detailed presentation of the feature extraction and classification process.

Figure 3 shows two representative examples of the classification results obtained by two different depth frames. The frames correspond to different body postures and the arms appear differently in each of the frames in terms of their projected position, real world position, and the shape/pattern of their arrangement. By observing the results, it is evident that in the produced label field the arms were clearly separated from the rest of the 3D scene with smooth outlines that extend from the shoulders to the wrists of the two arms. The locations of the segmented arms as well as the head of the driver were tracked over time in order to derive general driving patterns for each driver and the results are presented in detail in the *Results and Analysis* section.

Figure 3 goes here

2.2.2. Qualitative Analysis

The qualitative analysis looked beyond driver body posture and investigated driver behavior while performing these maneuvers. More specifically, the qualitative analysis process focused on examining drivers' body movements while performing various maneuvers and comparing the results with those from the quantitative analysis.

The following body movements were recorded as a part of the analysis:

- Head movements,
- Upper body movements, and
- Non-driving-related arm movements (e.g., adjusting the glasses, drinking water, adjusting the seatbelt).

The following assumptions were made for the data analysis:

- Head movements were recorded if the movement resulted in the driver completely losing sight of the road ahead. Primarily, a head movement to check the blind spot on either the left or right side of the driver was termed significant.
- Short and multiple head movements during a maneuver were considered significant and were recorded since they were believed to reduce the visibility of the road section ahead.
- Short and single head movements with duration of less than 0.15 seconds, involving

checking the side view mirror were ignored, since these were assumed to not cause significant loss of sight of the road.

- Minor movements such as head nodding or adjusting position on the seat that did not result in the driver losing the sight of the road were ignored.
- Turning of the upper body (shoulders) was termed significant and was recorded as a valid body movement. This movement usually supplemented the head movement of the participants, for example when checking their blind spot.

The start and end time of every movement was recorded along with the frequency of each movement during the corresponding time interval. The data were recorded for all available lane changing and merging maneuvers, as these were defined previously.

In addition, the time duration that drivers were not looking ahead while performing a lane change or a merging maneuver was examined. During these situations, the driver may be looking towards the side mirror or the rear-view mirror, checking the blind spots or looking at something non-driving related. Investigating the "eyes off the road" could reveal potential unsafe situations, especially if this is for a long period of time. To perform this analysis, the in-vehicle videos and the eye-tracking videos were used. The in-vehicle videos provided information on the vehicle environment while performing the maneuver, and the eye-tracking videos captured what the drivers were actually looking at. An example of the field of view of the eye-tracking camera and drivers' gaze is shown in Figure 4. In this study it was assumed that if an object appeared in the driver's horizon and it was close to the gaze crosshair, then the driver became immediately aware of that object. It was also assumed that, as shown in Figure 4, the crosshair did not necessarily have to coincide with the object in order for the driver to identify it.

3. RESULTS AND ANALYSIS

This section presents the analysis and results of the quantitative and the qualitative approach.

3.1 Quantitative Analysis

The quantitative analysis was performed using the depth sequences recorded from the Kinect sensor. In total, 523 depth video sequences of 27 drivers performing 236 merges and 287 lane changes were used in this analysis. The total number of frames contained in this dataset exceeds 300,000. Each depth frame sequence was segmented using the classification algorithm described earlier. Three different classifiers were trained using a manually defined training set in order to segment the regions of the left arm, right arm, and head of each driver respectively. The classifiers were modeled using standard multi-layer neural networks with 3 layers that consisted of 10 neurons in the input and intermediate layer, and two neurons in the output layer. The network was trained by performing least squares fitting to manually annotated depth frames for each driver in the dataset using the Levenberg-Marquardt optimization method (Lawson, 1974) and the back-propagation algorithm (Haykin, 2009).

The classification results were tracked across frames by computing the magnitude of the motion observed in each respective region. More specifically, the magnitude of the motion of the arms was computed by calculating $\Delta x^2+\Delta y^2+\Delta d^2$ of the 10 right-most pixels of each segmented region. The choice of these pixels approximated well the regions of the corresponding wrists; hence their tracking was considered a good descriptor of the arm activity during maneuvers. In order to enhance the robustness of the calculations a region of 10 pixels was used, and the average location (x, y, d) was computed from these pixels. Similarly, the average location was computed from the pixels of the entire head region computed from the classifier.

The total magnitude of the motion was computed individually for the left arm, right arm, and head for each of the 523 depth video sequences. The average and standard deviation of the magnitude of motion of each driver was calculated and plotted in Figure 5.

Figure 5 goes here

In Figure 5, the x-axis corresponds to different drivers in the dataset, and the y-axis corresponds to the calculated magnitude of motion. The average and standard deviation are shown as color-coded bar plots in blue, green, and red, for the head, right arm, and left arm respectively. By observing the plots, it is evident that the calculated motion of the region of the head was significantly smaller than the magnitude of motion observed in the arm regions, which was expected.

Another observation is that the right arm is more active than the left arm in the majority of the drivers, because the average magnitude (shown in dots) is slightly higher in the case of right arm (green dot) compared to the left arm (red dot). This result was also anticipated, as there is more physical space for right arm movements. However, it should be noted that right arm movement could correspond to non-driving related action, which was not separated from the statistical calculation. The standard deviation of the magnitude of motion was also notably larger for the region of the right arm that indicates inconsistent pattern across video sequences.

Similar analysis can be performed across various groups of drivers in our dataset. Figure 6 shows the statistics (average and standard deviation) of the motion magnitude of female drivers (11 subjects), male drivers (16 subjects), 20 year old drivers or younger (7 subjects), drivers between 20 and 30 years of age (13 subjects), and 30 year old drivers or older (6 subjects).

Figure 6 goes here

In general, minor variations were observed across the different groups of drivers. For instance, the average motion of the head was smaller in the female subjects compared to the male drivers, which could indicate either that more male subjects moved their head during maneuvers or that, in general, head motions were more frequent in the male drivers. On the other hand the subjects in the middle age group had slightly more intense arm motions compared to the younger or older subjects, which could either indicate that their driving pattern was more intense or that in general they moved their arms and especially the right one more frequently during the maneuvers.

Table 1 goes here

Since the above remarks were made by observing relatively small variations in the statistical results from Figures 5 and 6, we performed analysis of variances (ANOVA) in order to assess the weakness/strength of these observations. Standard f-tests were performed between various types of observations by using equal number of samples in each group. The probability *p* for the null hypothesis that the means of the groups are equal was calculated and reported in the last column of table 1, in addition to the corresponding f-values and the number of samples (N) in each group. The reported results indicate that the collected samples from the magnitude of the motion of the head, left arm and right arm are independent observations since the corresponding p values are below 5%. Furthermore the similarity between the motion data collected from the first and the third age group compared to the second age group appears strong (>92%). The comparison between the data collected from the male and female population was also assessed.

3.2 Qualitative Analysis

Concerning the qualitative analysis of the body movements, it was found that, compared to upper-body movements, head movements were the most predominant type of body movements while driving. All study drivers had a tendency to look over their shoulders and check the rear-view/side view mirrors to ensure whether it was safe or not to complete the merging or lane changing maneuvers. Also, most of the drivers used a substantial portion of their upper body (i.e., shoulders) when making the driving maneuvers, especially at the very instance of merging and lane changing. Non-driving related maneuvers included instances when the driver released his/her hand from the steering to perform non-driving related tasks such as: adjusting their glasses, drinking water, adjusting their seatbelt, hand gesturing when talking to other passengers in the vehicle, scratching their nose/hand, etc. The non-driving movements are not associated with distracted driving instances unlike talking on the cellphone, adjusting the radio, etc., therefore; these movements did not result in drivers taking their eyes off the road.

The start and end of every movement was recorded along with the frequency of each movement during the corresponding time interval. Table 2shows a sample for the qualitative analysis for merging maneuvers on SB Newberry Road for Driver ID 110. Data were recorded for every round completed by the driver. The start and the end times of the maneuver correspond to the frame on the Kinect player. To determine the total duration of the maneuver in seconds, the difference between the start and end frame was divided by 25 (25 frames per second). The frequency denotes the number of relevant body movements that the driver made during the corresponding time interval.

Table 2 goes here

After analyzing a total of 235 videos of merging maneuvers and 335 videos of lane changing maneuvers, the following generalized conclusions can be drawn:

- Head movements were the most predominant type of body movement when driving. In the majority of the cases head movements accounted for checking the blind spot before performing the required maneuvers.

- The average duration of a head movement was about 4 seconds for the merging maneuvers with a frequency of 2 movements, consistent for all three merge junctions.
- Interestingly, the average duration of head movement for a lane changing maneuver on the freeway was approximately 3.75 seconds with a frequency of 1.8 movements and the average duration of head movement for a lane changing maneuver on the arterial was approximately 2.3 seconds with a frequency of 1.1, consistent for both study locations. Therefore, drivers were more careful when making the lane changing maneuvers on the freeways than the arterials.
- Upper-body movements were observed only when a driver required checking the blind spot. Although, it was observed that the upper body movements were relative to each driver and varied from driver to driver, it cannot be concluded from the qualitative analysis that checking the blind spot involved the use of upper body movements for all the drivers. At the three merge junctions, the average duration of an upper body movement was 4.5 seconds while during lane changes, the average duration was 3.5 seconds.
- Non-driving related movements did not appear to impact the drivers' safety in this study. None of the drivers engaged in texting or using their cellphones during the course of the study, behaviors that often are associated with hampering of the drivers' vision of the road and diminished safety. The most common non-driving related movement was the drivers' tendency to use their hands for gesturing when talking to the vehicle occupants. Coincidentally the average duration of the non-driving related movements at the three merge junctions and at the two lane changing locations was 3.12 seconds. Thus, it could be suggested that non-driving related body movements are not correlated to the type of

the driving maneuvers involved and were performed by the drivers irrespective of the two maneuvers, i.e. merging and lane changing.

Lastly, during each lane changing maneuver performed, we investigated the amount of time that drivers' eyes were off the road when they performed that maneuver. This assessment focused on lane changing events, because in these situations the relationship with the lead vehicle is more prominent. In merging maneuvers a leading vehicle was typically absent from our database; thus, the need to evaluate the duration of time that the participants' eyes were off the road was minimal.

Data from 13 participants were used for the analysis of "eyes off the road" data during lane changing events. Each lane changing maneuver performed by the participants was looked thoroughly, and the times where the drivers were not looking at the leading vehicle were recorded. 3 presents the results of the analysis, along with some demographic information (gender and age) of the 13 study participants.

Table 3 goes here

As it can be seen in Table 3, the average time duration of eyes off the road ranges from 2 to 4 seconds and is similar for the study participants. This duration is also consistent with the average duration of head movement discussed earlier. However, the maximum duration where drivers were not looking at the vehicle in front differs significantly. Such situations could hide unsafe conditions during the lane changing maneuver. Unfortunately, the small variation between the ages of the study participants and the limited sample size available do not allow for an in depth analysis of the impact of age or gender on the likelihood of drivers taking their eyes off the road for long periods.

DISCUSSION AND CONCLUSIONS

The exact body posture of drivers while performing various maneuvers may reveal significant information with respect to what drivers actually look at and how safely they manage to interact with their environment. In addition, being able to understand and monitor driver behavior inside the vehicle has numerous applications regarding traffic safety. Although past research has looked at video-based systems for monitoring head movement and facial expressions, this research extended the state of the art by developing and testing an algorithm that tracked the position of drivers' arms and head with acceptable accuracy, assuming that an adequate training dataset is available. In addition, the results of the quantitative analysis showed that, drivers tend to move their right hands more often. Variations in the head/arm movements between female and male drivers, as well as drivers of different age groups were observed.

Apart from the quantitative analysis, investigation of the entire driver posture by using additional video sources and considering traffic conditions and the surrounding environment, provided some interesting conclusions. The average head movement duration during merging maneuvers was 4 seconds, whereas the corresponding duration for a freeway lane change was 3.75 seconds. On arterials, the average duration was found to be significantly less, at 2.3 seconds. This indicates that drivers are more cautious when driving on the freeway. Upper-body movements were observed only when a driver required checking the blind spot, and their average durations for merging and lane changing were 4.5 and 3.5 seconds, respectively. Lastly, non-driving related arm movements were not related to secondary activities, such as texting, talking on the phone, etc., therefore, there is no evidence from the study data that these negatively impact safety.

The qualitative study complements the results obtained from the quantitative approach, since the former looked at arm movements that are not related to driving and, therefore, capture instances where hands were off the wheel, while this was not possible to extract from the latter. In addition, the motion magnitude can only be inferred from the quantitative analysis, and not from visual inspections of the available videos.

It should be noted that, the research findings are very much related to the prevailing geometric and environmental conditions. It is anticipated that, in the event of adverse weather or complicated geometric features, driver behavior, body posture, head and arm movements and eye gaze will be different. Future research should focus on quantifying the effect of geometric and environmental conditions on driver body posture.

As a future step, the proposed algorithm could be further used to predict the intentions of the driver, identify potential hazardous conditions in vehicle's cabin and warn the driver accordingly through an advanced driver assistance system. For example, by monitoring drivers' body and arm movement through the developed algorithm, it is possible to develop an "average movement profile" for an individual driver, as a function of traffic, geometric, and environmental considerations, and use this profile to identify variations (i.e., due to distraction or other reason) in real time and provide feedback to the driver. For such applications, immediate feedback should be produced, and the computational complexity of the data processing algorithms should be such that real-time execution is feasible.

ACKNOWLEDGMENTS

This work was sponsored by US DOT/RITA through STRIDE (Project 2013-051S), the Alabama Department of Transportation (ALDOT) and the Florida Department of Transportation (FDOT). The support of the study sponsors is greatly appreciated. The authors also thank Mr. G. Yengera from the Indian Institute of Technology (BHU), Varanasi, India, for assisting with the neural network implementation.

REFERENCES

Barmpoutis A. (2013) "Tensor body: Real-Time Reconstruction of the Human Body and Avatar Synthesis from RGB-D". IEEE Transactions on Cybernetics 43(5), pp. 1347–1356.

Barmpoutis, A., Yengera, G., Hosuri, S. R., Islam, M. M., Kondyli, A., Patil, S. S., Sisiopiku, V. P., Zhang, L., and Zhao, L. (2015) "SHREC'15 Track: Driver's Motion Depth Database" , 3D Shape Retrieval Contest - Digital Worlds Institute Technical Report, pp. 1-7.

Campbell, B. N., Smith, J. D., Najm, W. G. (2003) "Examination of Crash Contributing Factors Using National Crash Databases", Report Number DOT-VNTSC-NHTSA-02-07, NHTSA, US DOT.

Cheng, S., and Trivedi, M. M. (2006) "Turn-Intent Analysis Using Body Pose for Intelligent Driver Assistance", IEEE Pervasive Computing, Vol. 5(4), pp. 28-37.

Dingus, T. A., Klauer, S. G., Neale, V. L., Petersen, A., Lee, S. E., Sudweeks, J., Perez, M. A., Hankey, J., Ramsey, D., Gupta, S., Bucher, C., Doerzaph, Z. R., Jermeland, J., and Knipling, R.R., (2006) "The 100-Car Naturalistic Driving Study, Phase II – Results of the 100-Car Field Experiment". Report No. DOT HS 810 593.

Haykin S. S. (2009) "Neural Networks and Learning Machines", Third ed. Pearson Education Upper Saddle River.

Hendricks, D. L., Fell, J. C., Freedman, M. (2001) "The Relative Frequency of Unsafe Driving Acts in Serious Traffic Crashes". Summary Technical Report for NHTSA, US DOT.

Kondyli, A., Sisiopiku, V. P., and Barmpoutis, A. (2013) "A 3D Experimental Framework for Exploring Drivers' Body Activity Using Infrared Depth Sensors". In Proceedings of the 2nd International Conference on Connected Vehicles & Expo (ICCVE 2013), Las Vegas, NV.

Lawson C., and Hanson R. (1974) "Solving Least Squares Problems". Prentice-Hall.

McCall, J., Wipf, D., Trivedi, M. M., and Rao, B. (2007) "Lane Change Intent Analysis Using Robust Operators and Sparse Bayesian Learning", IEEE Transactions on Intelligent Transportation Systems, Vol. 8(3).

Peden, M., Scurfield, R., Sleet, D., Mohan, D., Hyder, A. A., Jarawan, E., and Mathers, C. (2004) "World Report on Road Traffic Injury Prevention: Summary". In World Health Organization, Geneva, Switzerland.

Shaout, A., Colella, D., and Awad, S. (2011) "Advanced Driver Assistance Systems - Past, Present and Future". In 7th International Computer Engineering Conference (ICENCO), pp. 72-82.

Tijerina, L., Garrott, W. R., Stoltzfus, D., and Parmer, E. (2005) "Eye Glance Behavior of Van and Passenger Car Drivers during Lane Change Decision Phase". Transportation Research Record: Journal of the Transportation Research Board, 1937, Transportation Research Board of the National Academies, Washington, D.C., pp. 37-43.

Tran, C., and Trivedi, M. M. (2009) "Driver Assistance for Keeping Hands on the Wheel and Eyes on the Road", IEEE International Conference on Vehicular Electronics and Safety.

Trivedi, M.M., Gandhi, T., and McCall, J. (2007) "Looking-In and Looking-Out of a Vehicle: Computer-Vision-Based Enhanced Vehicle Safety". IEEE Transactions on Intelligent Transportation Systems, Vol. 8 (1), pp. 108-120.

US Department of Transportation, (2011) "USDOT Connected Vehicle Research Program Vehicle-to-Vehicle Safety Application Research Plan". Report No. DOT HS 811 373.

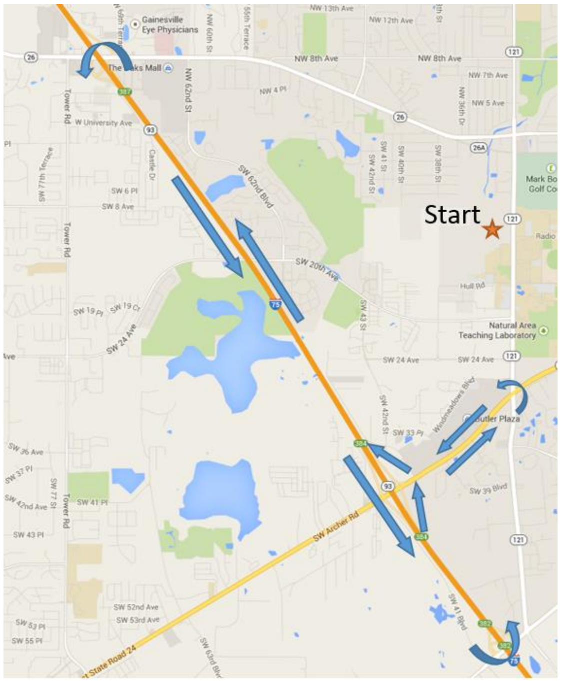

Figure 1: Data collection site

Figure 2: Example of a depth frame from the study dataset, visualized here as a 3D surface using computer graphics shading.

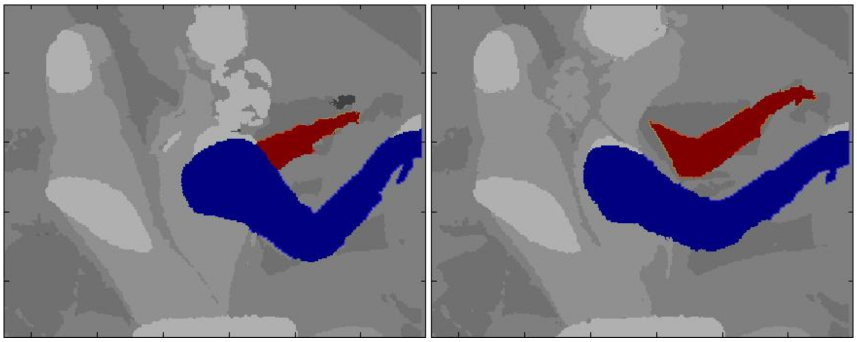

Figure 3: Example of the classification result for the left and right arm in two different depth frames.

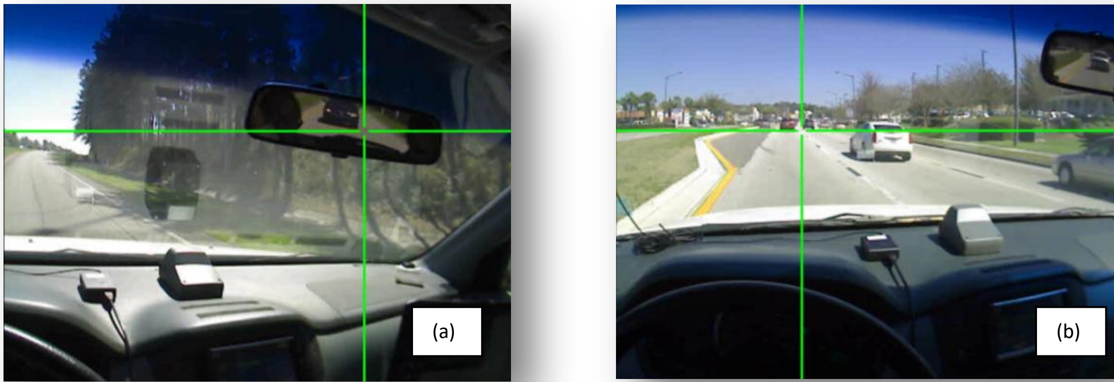

Figure 4: Field of view from the eye-tracking camera and gaze crosshair capturing (a) following vehicle through the rear-view mirror and (b) leading vehicles.

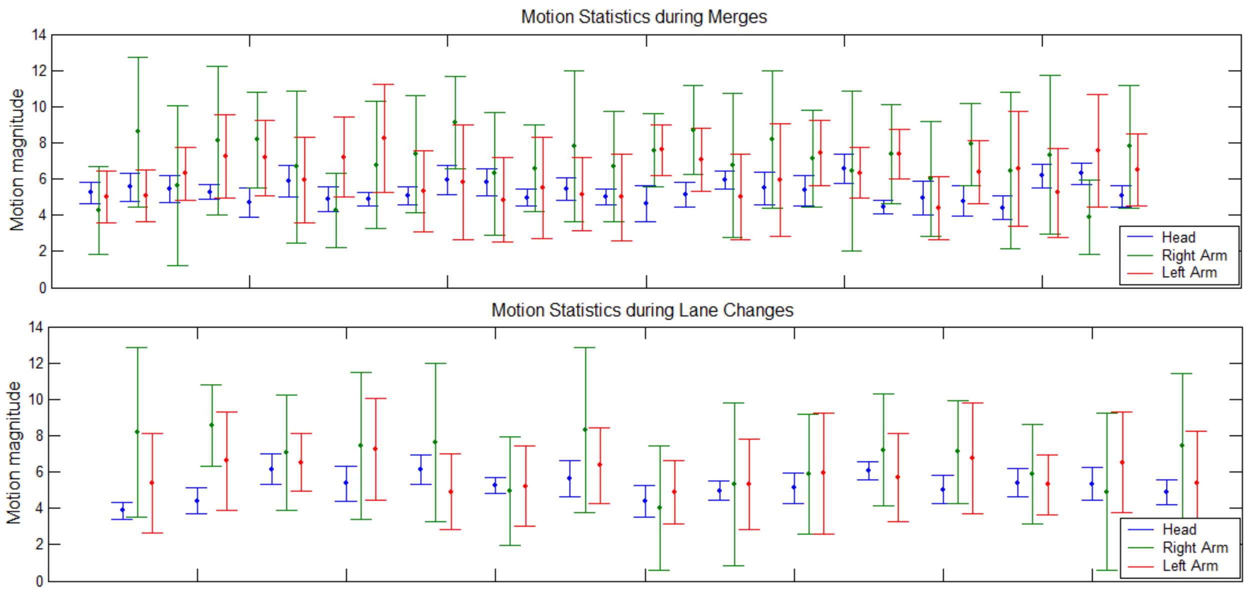

Figure 5: Plot of the statistics (average and standard deviation) of the magnitude of the motion observed in the regions of the left and right arms and head of each driver. The results are separately reported for merging and lane changing maneuvers.

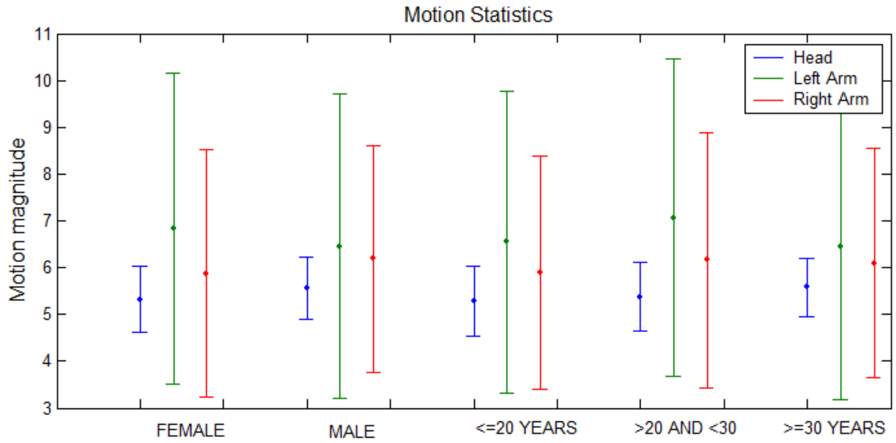

Figure 6: Plot of the statistics of the magnitude of the motion from various groups of drivers based on gender or age.

Table 1. ANOVA results of the merging maneuver data from various groups of drivers

Groups of observations	**N**	**F**	***p >F***
Head, Right Arm, Left Arm	27	12.1543	.000025417
Right Arm vs. Left Arm	27	4.0409	0.0496
Male vs. Female Head	11	2.0906	0.1637
Male vs. Female Arms	11	0.8562	0.3658
3 Age Groups Arms	6	1.6404	0.2268
Age Group 1 vs Age Group 2 Arms	6	2.0164	0.1860
Age Group 2 vs Age Group 3 Arms	6	2.3776	0.1541
Age Group 1 vs Age Group 3 Arms	6	0.0096	0.9239

Table 2. Sample Qualitative Data Analysis for Merging Maneuvers Performed by Driver ID 110 on SB Newberry Rd.

Driver ID	**Type of Body Movement**											
	Head				**Upper body**				**Non-driving related arm movement**			
	Start time	**End time**	**Duration (s)**	**Frequency**	**Start time**	**End time**	**Duration (s)**	**Frequency**	**Start time**	**End time**	**Duration (s)**	**Frequency**
	422	440	0.72	1								
	500	515	0.60	1								
	639	852	8.52	3	639	688	1.96	1				
	34	53	0.76	1								
	461	490	1.16	1								
110	514	819	12.20	5	539	591	2.08	1				
	365	397	1.28	1					793	835	1.68	1
	500	525	1.00	1								
	598	752	6.16	4								
	793	835	1.68	1								
	932	953	0.84	1								

Table 3. Summary of Eyes off the Road Time when Performing Lane Changing Maneuvers Findings

Driver ID	# of Lane Changes	Average Duration	Median Duration	Min Duration	Max Duration	Gender	Age
103	12	0:00:04	0:00:04	0:00:01	0:00:07	F	35
104	8	0:00:04	0:00:03	0:00:01	0:00:07	F	21
106	7	0:00:03	0:00:02	0:00:01	0:00:07	M	16
109	21	0:00:03	0:00:03	0:00:01	0:00:08	M	24
111	25	0:00:04	0:00:03	0:00:01	0:00:15	M	30
118	7	0:00:02	0:00:02	0:00:01	0:00:03	M	20
119	4	0:00:04	0:00:03	0:00:01	0:00:07	F	25
120	20	0:00:04	0:00:03	0:00:01	0:00:08	M	20
121	29	0:00:03	0:00:02	0:00:01	0:00:08	F	22
122	23	0:00:03	0:00:03	0:00:01	0:00:09	M	54
124	13	0:00:03	0:00:03	0:00:01	0:00:10	M	21
125	24	0:00:03	0:00:02	0:00:01	0:00:09	M	21
126	17	0:00:03	0:00:03	0:00:01	0:00:06	F	25