A 3D experimental framework for exploring drivers' body activity using infrared depth sensors

Alexandra Kondyli
Department of Civil and Coastal Engineering
University of Florida
Gainesville, Florida 32611–6580
Email: azk133@ufl.edu

Virginia Sisiopiku
Department of Civil, Construction, and Environmental Engineering
University of Alabama
Birmingham, Alabama 35294-4440
Email: vsisiopi@uab.edu

Angelos Barmpoutis
Digital Worlds Institute
University of Florida
Gainesville, Florida 32611-5800
Email: angelos@digitalworlds.ufl.edu

***Abstract*—A significant amount of research has been involved with the development of advanced driver-assistance systems. Such systems typically include radars, laser or video sensors that detect the vehicle trajectory and warn for an imminent lane departure, or sense the front vehicle's speed and apply the brakes of the following vehicle to maintain safe distance headways (i.e., collision avoidance system). However, most of these systems rely on the subject vehicle and surrounding vehicles' position and do not explicitly consider the driver's actions during the driving task. In addition safety research has focused on eye tracking as a means of capturing driver's attention, fatigue, or drowsiness; however, the body posture has not been investigated in depth. This paper presents a novel approach for studying the actual movements of drivers inside the vehicle, when performing specific maneuver types such as lane changing and merging. This information can be useful for identifying specific body movements that may hide potentially unsafe situations. A pilot study was conducted along a freeway and arterial segment, where the 3D shapes of selected participants were constructed with the use of a low-cost infrared depth sensor (Microsoft Kinect) while merging and changing lanes. The analysis of the 3D shapes shows that there are important differences between participants when performing similar driving maneuvers. The preliminary results of this pilot research set the basis for implementing the proposed methodological framework for conducting full-scale experiments with a variety of participants, and exploring differences due to driver behavior attributes, such as age, gender and driving experience.**

I. 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 [1]. According to the World Health Organization (WHO), annually there are over 1.2 million fatalities and over 20 million serious injuries worldwide. In the US, the 100-car naturalistic study sponsored by the National Highway Traffic Safety Administration (NHTSA) concluded that driver inattention is the cause of about 80 percent of crashes and 65 percent of near crashes [2]; and therefore, these can be avoidable. A lot of attention has been drawn lately to US Department of Transportation (USDOT) connected-vehicle research program, which uses a mixture of technologies such as advanced wireless communications, onboard computer processing, advanced vehicle-sensors, GPS navigation, and smart infrastructure, to identify and warn the drivers on imminent road hazards [3]. The program includes vehicle-to-vehicle and vehicle-to-infrastructure communication research activities. The vehicle-to-vehicle communication is related to the exchange of data (e.g., speed, acceleration, heading angle, etc.) over wireless network that provide information on surrounding vehicles status and allows for performing calculations and issue driver warnings to avoid crashes. The communication option is based on the Dedicated Short Range Communications (DSRC). Although the development of the communication component of this program is not complete to date, a number of crash avoidance systems (e.g., blind spot and lane changing warning, forward collision warning, etc.) has been established so far.

Additional advanced (or intelligent) driver assistance systems (ADAS) designed to provide added traffic safety are already in place [4]. These systems typically 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. These monitoring systems are capable of tracking driver's inattention and drowsiness using LED sensors to monitor eye movement.

In vision-based systems that involve understanding driver intentions and actions (e.g., inattention or distraction states), research studies focus primarily on tracking of the head and the face of the driver e.g., [5], [6] and constructing 3D space images using the geometry of the face [7], [8], [9]. In addition, several researchers, e.g., [10], [11], [12], [13] analyzed head pose and gaze for identifying and predicting driver's intent to change lanes and perform a maneuver. Apart from tracking head and facial poses, research has also studied the hand position and grasp in conjunction with head monitoring for lane change intent analysis and prediction [14] or for driver distraction monitoring [15]. Another study [16] presented a system for tracking the 3D body movement combined with head pose tracking system. 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 is promising, their results to date are limited. [17] expanded their work to investigating drivers' foot behavior using video-based analysis in conjunction with

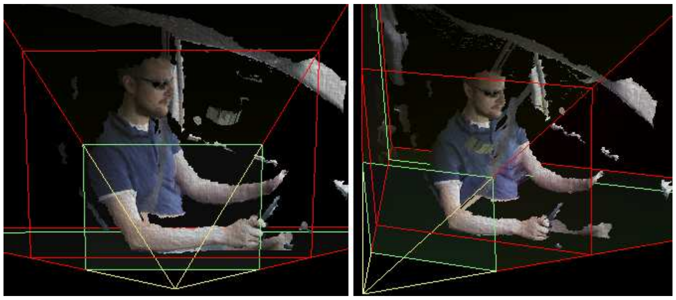

Fig. 1. Two 3D views of the same frame from the recorded dataset. The video and depth frames are presented as a sequence of textured 3D frames. The field of view of the depth camera is also shown as a trapezoid.

Fig. 2. Left: visualization of a depth frame. Right: The corresponding mask with enhanced boundaries between objects, computed using our framework.

pedal sensor measurements. They presented a prediction model for braking and acceleration modes and concluded that the foot behavior depends greatly on the driver type. However, several limitations were identified, particularly with respect to the computational effort of foot tracking, which may result in delayed predictions that can be critical.

In summary, although a significant amount of research has been involved with the development of advanced driver-assistance systems, 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, recent research has focused on eye/facial tracking as a means of capturing driver's attention, fatigue, or drowsiness. To date, limited research has been involved with investigating the upper-body posture of drivers when performing a maneuver as well as different postures between different drivers, which may also reveal behaviors that contribute to unsafe driving conditions.

Furthermore, several of the aforementioned problems lie on the fact that the existing vision-based techniques employ 2D image computer vision algorithms that may lead to inaccuracies when computing 3D data due to lack of the depicted information [18]. It has been shown that many traditional computer vision problems can be solved more efficiently and/or accurately using depth cameras in conjunction with regular video [19]. When it comes to pose estimation [20] or 3D reconstruction of the human body [21], [22], it has been shown that infrared depth sensors can estimate the shape characteristics of the human body in real-time [23], which has numerous applications in various research areas ranging from human-computer interaction to monitoring obesity [24].

The main objective of this paper is to investigate how driver posture and activity during the driving task can be obtained and analyzed in real-time using a low-cost infrared depth sensor. The findings of this research will assist in identifying the necessary tools for exploring the correlation between potentially unsafe driving conditions and body posture as a function of specific driver characteristics and attributes. These findings could lead to enhancing advanced driver-assistance systems by identifying specific body activity associated with unsafe conditions under different maneuvers.

The contributions in this paper are threefold: 1) We introduce the use of infrared depth cameras as an intelligent sensor for monitoring the driver's body activity. 2) We propose a basic framework for acquiring, segmenting, analyzing, and visualizing the recorded sequences of depth frames. 3) We present the efficacy of the proposed methods using several experimental results from a pilot study.

II. Methods

Each data frame captured by a digital depth sensor is a two dimensional array of depth values (i.e., distance between the sensor and objects). Similarly, a collection of frames is a three dimensional array that can be represented as $\mathbf{D} \in \mathbb{R}^{W \times H \times N}$, where N denotes the total number of recorded frames, and W and H denote the number of pixels across the width and height of the depth frame respectively. The depth value in a particular pixel with coordinates (i, j) on frame i is denoted by $D_{i,j,t} \in \mathbb{R}^{+}$. In practice, each depth camera has a specific range of operation, which restricts accordingly the range of the recorded values (see depicted field of view in Figure. 1).

The depth frames can be equivalently expressed as quadratic meshes given by $X_{i,j,t} = (i - i_c)D_{i,j,t}f^{-1}$, $Y_{i,j,t} = (j - j_c)D_{i,j,t}f^{-1}$, and $Z_{i,j,t} = D_{i,j,t}$, where (i_c, j_c) denote the coordinates of the central pixel in the depth frame, and f is the focal length of the depth camera. One of the advantages of the quadratic mesh representation of the depth frames is that they can be easily visualized using virtual lighting, shading, perspective and point of view using standard computer graphics techniques [18]. Figure 2(left plate) shows the quadratic mesh of a captured depth frame from our pilot study. The 3D shape of the body of the driver and part of the vehicles' cabin have been clearly captured in the depth frame. Optionally, the color information from a video frame can be applied as a texture to the quadratic mesh of the depth frame. Two examples of such visualization is shown in Fig. 1.

The segmentation of the depth frames is a necessary preprocessing step for analyzing the activities of the human body. The process of image segmentation is a well-studied computer vision problem [25], which may be inaccurate when adjacent regions have similar color patterns, and there is no clear boundary between them. In our proposed framework, the information captured in the depth frames is enough for estimating accurately the outlines or boundaries between critical regions in the field of view, such as the driver's arms, as follows: For each depth frame, a binary mask is computer by evaluating the following two conditions for every pixel x, y and frame t

- $max_{x,y \in N(i,j)} |D_{i,j,t} - D_{x,y,t}| < threshold_{dz}$
- $min_{s \in N(t)} D_{i,j,s} > float_{err}$,

Fig. 3. Example of the skeleton model that was fit to an arbitrary frame of the depth sequence. The 3D coordinates of the shoulders, neck, and head, were employed for real-time segmentation of the body activity. The lower joints (elbows, wrists, and hands) were ignored due to inaccuracies in their fitting. Instead we use the proposed arm segmentation method (Fig. 4).

Fig. 4. Two examples of the proposed arm segmentation. Both arms can be clearly segmented from the rest of the depth frame even when one arm is occluded or partially visible from the depth camera (right).

where $N(t)$ and $N(i,j)$ denote 1D and 2D sets of integers in the neighbor of the input t, and i,j respectively, and $threshold_{dz}$, and $float_{err}$ are two predefined constants. Each pixel for which both conditions are true is considered part of the depicted object in contrast to the rest of the pixels that belong to the boundary between regions or to an empty space. The role of the first condition is to segment together pixels with similar depth values, and the second condition ignores pixels with: a) depth values in the range of a computer precision error and/or b) inconsistent depth estimation across neighboring frames. Figure 2 shows an example of a computed mask with clear outlines around the depicted objects.

The computed mask is then used to segment the driver's arms by counting and labeling all independent regions within the volume that is defined between the driver's torso and the steering wheel. In our experiments, the position of the driver's torso was estimated using the skeleton fitting process provided in the Microsoft's Kinect Software Development Kit [26]. The fitted skeletal model consists of the 3D coordinates of 10 major joints in the upper part of the body (head, neck, shoulders, elbows, wrists, hands). Figure 3 shows an example of the fitted skeleton in one frame from our dataset. Based on the skeleton fitting results from our entire dataset, only the upper joints (head, neck, shoulders) were consistently estimated accross frames, while the rest of the joints were misfit in the majority of depth frames. In order to estimate the position of the driver's torso and more specifically the x-location of the plane which is parallel to the driver's chest and is perpendicular to the camera plane, we used the average of the x-coordinates of the head, neck and shoulders from the estimated skeletons.

After estimating the rectangular volume between the driver's torso and the steering wheel, we segmented the individual regions using the aforementioned mask, and the average x,y,z coordinates were computed from the pixels of each region. The regions with the highest y-value corresponded to the driver's arms, and the arm with the smallest z-value was the right arm (i.e. closest to the camera). Figure. 4 shows two examples of segmented depth frames. By observing the images, it is evident that the arms were accurately segmented independently of the relative position of the two arms. In our experiments we used the segmented regions of the arms as well as the upper joints of the tracked skeleton in order to track the motion of the arms and head of the driver while driving.

To track the body movements we estimated $\frac{\partial \mathbf{D}}{\partial t} \sim D_{i,j,t} - D_{i,j,t-1}$ for every i,j,t and then we computed the average of the negative values and the average of the positive values within each region. The magnitude of these two average values correspond to the directional magnitude of inward and outward motion with respect to the z-axis. The directional magnitude of motion is shown in several of our examples (Figs. 6, 7, 8).

Finally, global statistics were computed accross several depth frames in order to study the variations of such global quantities between different drivers. More specifically, the mean depth frame was computed as $M_{i,j} = \sum_t D_{i,j,t}$, and the standard deviation $S_{i,j} = \sqrt{\frac{1}{N}\sum_t (D_{i,j,t} - M_{i,j})^2}$, which can both be considered depth frames and therefore can be visualized similarly (see Figure 6).

The following sections present a description of a pilot study undertaken to collect field observations of drivers' 3D body shapes and several experimental results obtained using the proposed methods.

III. Driver Behavior Data Collection

The field data obtained for this study were collected along a 2.6 mi stretch of Interstate 75 (I-75) in the southbound (SB) and northbound (NB) directions, and a 0.7 mile long arterial segment (Newberry Road eastbound and westbound approaches) in Gainesville, FL. The freeway segment has three lanes per direction and the arterial segment has three through lanes per direction, several median openings, and includes a total of six signalized intersections. A schematic of the study sites is presented in Figure 5. The data collection effort took place on Sunday, September 1st 2013, between 10 am and noon. Traffic conditions were generally uncongested and free-flowing, especially on the freeway segment. Traffic on the arterial segment was light, although towards the end of the data collection effort the flows were considerably increased. For the purposes of this pilot study, four participants affiliated with the research team were asked to complete one route along the freeway and arterial segment. The participants performed two mandatory lane changes (i.e., merging onto the freeway) and several discretionary lane changes on the freeway and the arterial street. The entire duration of the experiment for each participant was approximately 20 minutes.

The real-time driver behavior data were acquired using the PrimeSense™depth sensor contained in the Microsoft Kinect™device. The device was connected (via a USB 2.0 port) to a 64-bit computer with Intel Core i5 (quad core)

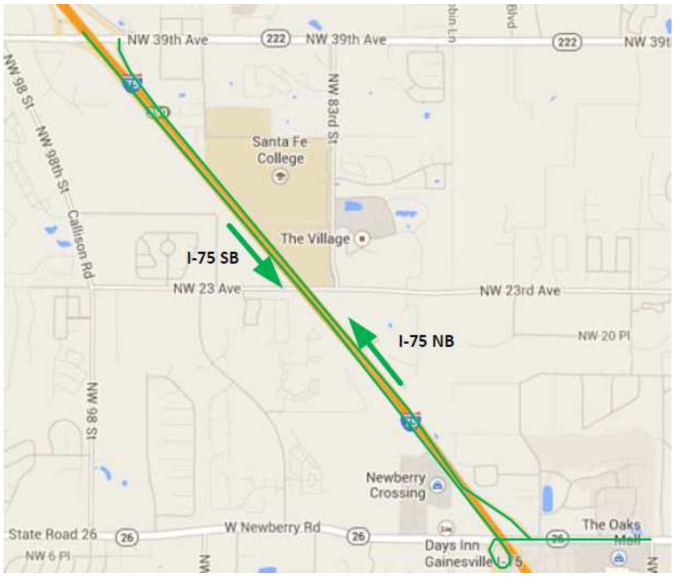

Fig. 5. Map of the route on the interstate I-75 followed in this pilot study.

CPU at 2.53GHz and 4GB RAM. The computer and the sensor were both powered using a 75 Watt car power inverter. The resolution of the depth camera was 320×240 pixels with horizontal field-of-view angle (FoV) angle of 57^o. The resolution of the video camera was 640×480 pixels with horizontal FoV of 62^o.

The range of the camera was calibrated so that it records depth values in the range from 0.5m to 3.0m, which is suitable for the limited space of the cabin of a typical passenger vehicle. The sensor was fixed on the front passenger's door, so that the driver is within the field of view of the depth and video cameras. Figure 1 shows the field of view of the depth camera. The green rectangle depicts the closest plane of sensing, which is located 0.5m in front of the sensor (shown as the tip of the yellow pyramid in the same figure).

IV. Discussion of Experimental Results

The video and depth sequences captured during our pilot study, were manually segmented into several fragments that correspond to the merging and exiting from the highway as well as changing lanes, right, and left turning in arterial streets. Each of the fragments was analyzed independently using the framework that was presented in Sec. II, and a comparative study was performed across the corresponding datasets from different participating drivers. The proposed framework was implemented in Java using the J4K open source Java library for Kinect that was originally presented in [23] and is available at http://www.digitalworlds.ufl.edu/angelos/lab/kinect.

Figure 6 shows the average and standard deviation of the depth sequence during merging (left) and exiting (right) a highway. The average depth value in each pixel forms a surface, which can be plotted in 3D using photorealistic shading to visually enhance the depicted depth information. The standard deviation of the depth values can either be presented as a surface or as a color map added to the average surface as shown in Fig. 6. In our plots, the intensity of the red color is proportional to the standard deviation of the depth values in the corresponding pixel. Large standard deviation values indicate wide range of motion at the corresponding pixels during the data sequence. As expected, an exit from a highway through a loop ramp is typically accompanied by a wide turn, which caused in the right image of Fig. 6 significant motion in the area of the arms.

During the merging maneuver, it can be observed that the motion of the arms and head, although significantly less, is still distinguishable and can provide important information of the participant's body posture while merging. For instance, the analysis of the mean and standard deviation might indicate that the specific participant made use of the side mirrors for completing the merging maneuver, instead of turning thoroughly the head and investigate potentially unsafe conditions. Fig. 6 also shows that even incremental variations of the body posture can be captured, which validates the proposed method. This type of variations may be significant when evaluating the variability of body movement across different driver types and under different driving situations.

Apart from the mean and standard deviation of the depth sequence, we can identify the exact direction of each movement and associated magnitude, as a function of the increase or decrease of the depth values. For example, Fig. 7 and Fig. 8 show the directional magnitude of the head motion and the arms motion respectively, for two of the subjects participated in the pilot study. The investigation of the magnitude of each movement may reveal interesting trends for each individual participant. First of all, it is possible to consider both movements of the arms and head in conjunction and not in isolation, contrary to previous studies that treat these two separately. Then, we can associate both movements with a specific maneuver (i.e., merging, lane changing, etc.) and construct a profile for each individual participant based on their typical behavior and movement activity. Such analysis will quantify differences in body postures between different driver types and could point out towards behaviors that may lead to potentially unsafe and even accident-prone driving conditions.

An example of such analysis is illustrated in Figs. 9 and 10 that show the directional magnitudes of the head and arms while merging for all four participants. The differences in the magnitude as well as duration of head and arms directional change is apparent in these figures. We further note the variability observed due to driver behavioral attributes and also due to traffic conditions. For instance, drivers 2 and 4 appear to have increased arm and head movement compared to drivers 1 and 3. In addition, driver 2 appears to have increased head activity at several instances (e.g., note the three peaks in the graph of Fig. 9), which may indicate increased alertness while merging, possibly due to the presence of a vehicle in the right-most lane. In addition, these two figures show that for some drivers, the head and arm movement is somewhat synchronized, although the arm movement is considerably more intense, as expected.

The depth sequence of each pixel can also be used for constructing a skeleton model, similar to that presented in Fig. 3. The 3D coordinates of the shoulders, neck, and head, were employed for real-time segmentation of the body activity. The movement of both arms was captured through the arm segmentation method illustrated in Fig. 4. This type of analysis is focused on a comprehensive investigation of the body

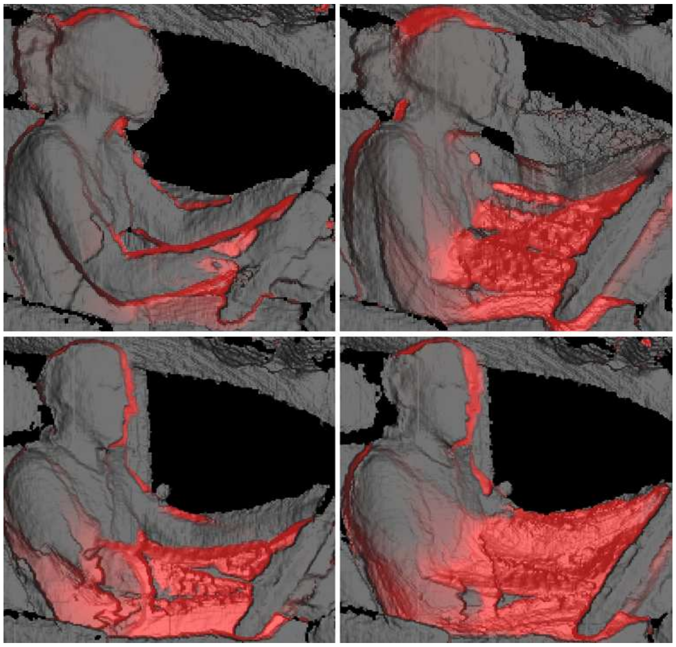

Fig. 6. Global statistics (mean and standard deviation across time) computed during merging (left) and exiting (right) from a highway for two different drivers (upper and lower row respectively). The mean is shown as the depth value within each pixel, and the st. dev. is shown as the intensity of red.

activity, rather than looking in isolation the various parts of the upper body.

It should be noted that the proposed analysis of the mean and standard deviation of the depth sequence shown in Fig. 6 as well as the directional magnitude of both head and arms motion provides significant insights related to the body posture and movements during various driving tasks, such as merging, changing lanes, as well as while undertaking secondary tasks such as texting, talking on the cell phone, eating, tuning the ratio, etc. In addition, the methodological framework described in this paper is capable of capturing variations across drivers, by examining differences of the mean and standard deviation of the depth sequence and the directional magnitude of motion for different driving maneuvers. These findings are useful for enhancing or developing an advanced drivers' assistance system that is able to detect driver motion and predict potentially unsafe conditions, and therefore, provide warning to the driver. This type of warning would complement existing surveillance systems typically installed to warn the driver for the surrounding traffic and the vehicle position.

V. Conclusion

A novel approach for assessing drivers' body movements inside a vehicle was introduced in this paper. The proposed method can be used for investigating how different driver types perform various maneuvers and which specific movements are associated with safe or unsafe driving conditions. A pilot study was conducted as a proof of concept, where four participants drove along a freeway and arterial route and performed a number of merging and lane changing maneuvers. The 3D shapes of the participants were constructed with the use of a low-cost infrared depth sensor for each maneuver performed. Several quantitative measures were evaluated as part of the preliminary analysis of the pilot study. Global statistics such as the mean and standard deviation as well as the directional movement of motion revealed significant differences for different maneuver types and among the participants. Contrary to current research, the proposed methodology may be used for studying the upper body posture and motion as a whole, instead of focusing on

Fig. 7. An example of a video frame detecting intense head motion. On the right the corresponding computed directional magnitude of motion is shown in red or blue for increase or decrease of the depth values.

Fig. 8. Another example with intense motion of the arms. The colors and images are presented using the same format as in Fig. 7.

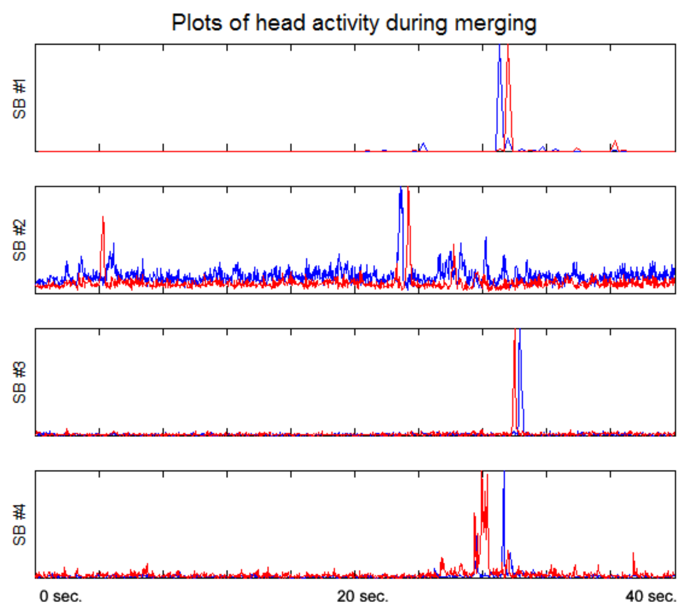

Fig. 9. Plot of the directional magnitude of the head motion during merging for 4 different drivers. The red and blue colors represent increase and decrease of the depth values respectively.

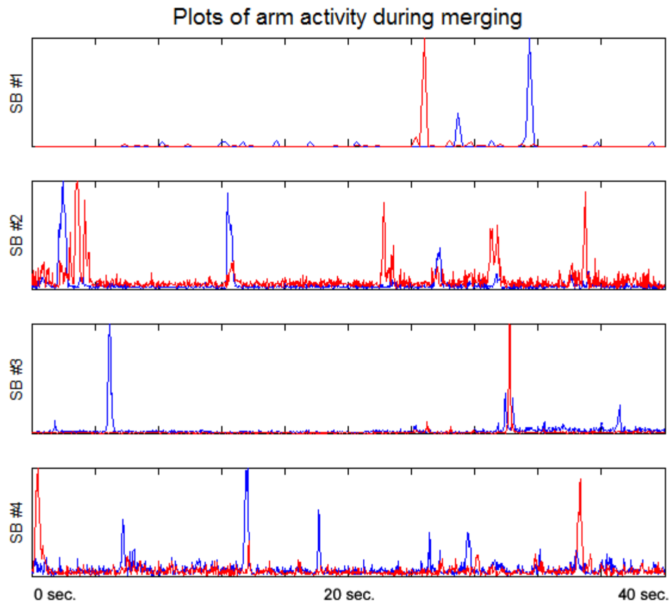

Fig. 10. Plot of the directional magnitude of the arm motion during merging. The format of the plots is the same as in Fig. 9.

individual parts of the body in isolation.

A future direction is to expand the implementation of the proposed methodological framework to additional drivers and investigate the relationship between potentially unsafe driving events and the actual driver body posture and movements when performing a driving maneuver (e.g., lane changing, merging) under different traffic and geometric configurations and when engaging with a secondary task by analyzing the movement of various drivers. We will also identify typical behaviors of specific driver groups (e.g., younger vs. older drivers, aggressive vs. conservative drivers, men vs. women), in naturalistic settings. Such information can be used for enhancing current driver training methods for targeted driver groups such as novice or elderly drivers. Lastly, it is recommended to develop a framework for constructing an in-vehicle driver-assistance system that takes into account the driver's body posture and movements rather than considering solely the vehicle position.

Acknowledgment

This project was in part funded by the grant award 2013-051S from STRIDE / US Department of Transportation. The authors would like to thank the anonymous volunteers who participated in this pilot study.

References

[1] M. Peden, R. Scurfield, D. Sleet, D. Mohan, A. Hyder, E. Jarawan, and M. Mathers, "World report on road traffic injury prevention," World Health Organization, Tech. Rep., 2004.

[2] T. A. Dingus, S. Klauer, V. L. Neale, A. Petersen, S. E. Lee, J. Sudweeks, M. A. Perez, J. Hankey, D. Ramsey, S. Gupta, C. Bucher, Z. R. Doerzaph, J. Jermeland, and R. Knipling, "The 100-car naturalistic driving study, phase ii results of the 100-car field experiment, report no. DOT HS 810 593," National Highway Traffic Safety Administration (NHTSA), Tech. Rep., 2006.

[3] "Connected vehicle research program vehicle-to-vehicle safety application research plan, dot hs 811 373," 2011.

[4] A. Shaout, D. Colella, and S. Awad, "Advanced driver assistance systems - past, present and future," in *Computer Engineering Conference (ICENCO), 2011 Seventh International*, 2011, pp. 72–82.

[5] K. Huang and M. Trivedi, "Robust real-time detection, tracking, and pose estimation of faces in video streams," in *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, vol. 3, 2004, pp. 965–968.

[6] E. Murphy-Chutorian and M. Trivedi, "Head pose estimation and augmented reality tracking: An integrated system and evaluation for monitoring driver awareness," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 11, no. 2, pp. 300–311, 2010.

[7] B. Braathen, M. Bartlett, G. Littewort-Ford, and J. Movellan, "3-d head pose estimation from video by nonlinear stochastic particle filtering," in *Proceedings of the 8th Joint Symposium on Neural Computation*, 2001.

[8] K. Huang, M. Trivedi, and T. Gandhi, "Driver's view and vehicle surround estimation using omnidirectional video stream," in *Intelligent Vehicles Symposium, 2003. Proceedings. IEEE*, 2003, pp. 444–449.

[9] J. Wu and M. Trivedi, "A two-stage head pose estimation framework and evaluation," *Pattern Recognition*, vol. 41, pp. 1138–1158, 2008.

[10] L. Tijerina, D. Stoltzfus, and E. Parmer, "Eye glance behavior of van and passenger car drivers during lane change decision phase," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1937, p. 3743, 2005.

[11] M. Trivedi, T. Gandhi, and J. McCall, "Looking-in and looking-out of a vehicle: Computer-vision-based enhanced vehicle safety," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 8, no. 1, pp. 108–120, 2007.

[12] J. McCall, M. Trivedi, D. Wipf, and B. Rao, "Lane change intent analysis using robust operators and sparse bayesian learning," in *Computer Vision and Pattern Recognition - Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on*, 2005, pp. 59–59.

[13] A. Doshi, B. Morris, and M. Trivedi, "On-road prediction of driver's intent with multimodal sensory cues," *Pervasive Computing, IEEE*, vol. 10, no. 3, pp. 22–34, 2011.

[14] S. Cheng and M. Trivedi, "Vision-based infotainment user determination by hand recognition for driver assistance," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 11, no. 3, pp. 759–764, 2010.

[15] C. Tran and M. Trivedi, "Driver assistance for "'keeping hands on the wheel and eyes on the road"'," in *Vehicular Electronics and Safety (ICVES), 2009 IEEE International Conference on*, 2009, pp. 97–101.

[16] ——, "Towards a vision-based system exploring 3d driver posture dynamics for driver assistance: Issues and possibilities," in *Intelligent Vehicles Symposium (IV), 2010 IEEE*, 2010, pp. 179–184.

[17] C. Tran, A. Doshi, and M. Trivedi, "Modeling and prediction of driver behavior by foot gesture analysis," *Computer Vision and Image Understanding*, vol. 116, pp. 435–445, 2012.

[18] O. Faugeras, *Three-Dimensional Computer Vision*. MIT press, 1993.

[19] J. Han, L. Shao, D. Xu, and J. Shotton, "Enhanced computer vision with Microsoft Kinect sensor: A review," *Cybernetics, IEEE Transactions on*, vol. 43, no. 5, p. in press, 2013.

[20] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, "Real-time human pose recognition in parts from single depth images," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, 2011, pp. 1297–1304.

[21] J. Tong, J. Zhou, L. Liu, Z. Pan, and H. Yan, "Scanning 3d full human bodies using kinects," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 18, no. 4, pp. 643–650, 2012.

[22] A. Weiss, D. Hirshberg, and M. Black, "Home 3d body scans from noisy image and range data," in *Computer Vision (ICCV), 2011 IEEE International Conference on*, 2011, pp. 1951–1958.

[23] A. Barmpoutis, "Tensor body: Real-time reconstruction of the human body and avatar synthesis from rgb-d," *IEEE Transactions on Cybernetics*, vol. 43, no. 5, pp. 1347–1356, 2013.

[24] ——, "Automated human avatar synthesis for obesity control using low-cost depth cameras," *Stud. Health Technol. Inform.*, vol. 184, pp. 36–42, 2013.

[25] D. Forsyth and J. Ponce, *Computer Vision: A modern approach*. Prentice Hall, 2003.

[26] "Microsoft Kinect SDK," *http://www.microsoft.com/enus/kinectforwindows/*.