Computer assisted analysis of drivers' body activity using a range camera

Alexandra Kondyli, Virginia P. Sisiopiku, Liangke Zhao, Angelos Barmpoutis, *Member, IEEE*

(Invited Paper)

***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. A pilot study was conducted along a freeway and arterial segment, where the 3D shapes of selected participants were constructed with the use of Microsoft Kinect range camera while merging and changing lanes. A 7-point human skeletal model was fit to the captured range data (depth frame sequences) using the proposed framework. The analysis of the captured 3D data showed 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, on-board 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 refers 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 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.) have been established so far.

Additional advanced driver assistance systems (ADAS) designed to provide added traffic safety are already in place [4]. These systems 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 condition. These monitoring systems for example, 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. Researchers in [17] expanded their work to investigating drivers' foot behavior using video-based analysis in conjunction with 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.

Accepted for publication in the IEEE Intelligent Transportation Systems Magazine, 2015.

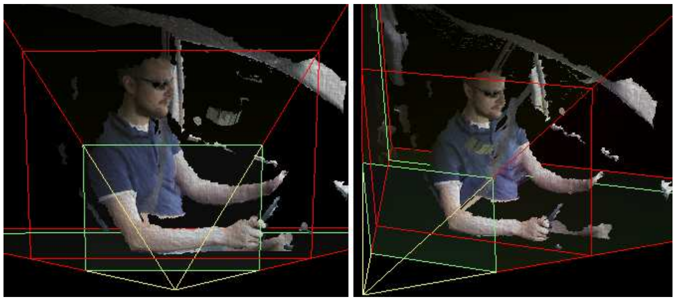

Fig. 1. Two 3D views of the same frame from the recorded dataset. The video and depth frames are presented as a sequence of textured 3D frames. The field of view of the depth camera is also shown as a trapezoid.

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 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 focused on 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 relay to 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. It has been shown that many traditional computer vision problems can be solved more efficiently and/or accurately using range cameras in conjunction with regular video [18]. When it comes to pose estimation [19] or 3D reconstruction of the human body [20], [21], it has been shown that depth sensors can estimate the shape characteristics of the human body in real-time [22], which has numerous applications in various research areas from human-computer interaction to rehabilitation [23] and monitoring obesity [24].

One popular area of application of human body tracking algorithms is electronic games. There are several examples in literature that report novel uses of body tracking technologies in games [25], or the development of novel algorithms for custom-made interaction using special-purpose partial body tracking [26]. An interactive game-based rehabilitation system using Kinect was presented in [25]. For a comprehensive literature review regarding the use of virtual reality and interactive games for rehabilitation the reader is referred to [27]. In most of these applications it has been shown that the existing body-tracking algorithms pose significant limitations such as constraints on the environmental setup, requirements regarding the pose of the users, the number of users being recognized, the number of 3D points tracked, etc. For example in [23] it has been shown that generic game-based depth-camera tracking algorithms fail in complex environments, when the human body is in close proximity with other objects or subjects in the field of view. As we also demonstrate in Sec. II, the same limitations apply to the case of vehicle's cabin [28]. In this paper, we overcome these issues by proposing a novel special-purpose body-tracking framework that focuses on detecting and tracking joints of the upper-body of the driver in a typical vehicle environment.

The main objective of this paper is to investigate how driver's posture and activity during the driving task can be obtained and analyzed in real-time using a low-cost range camera. The findings of this research will assist in identifying the necessary tools for exploring in the future the correlation between potentially unsafe driving conditions as a function of specific driver body postures and actions, such as talking to phone, texting while driving, or engaging in other non driving-related activities. These findings could lead to enhancing advanced driver-assistance systems by identifying specific body activity associated with unsafe conditions.

The contributions in this paper are three-fold: A) A new framework is proposed for studying research questions regarding the driver's body posture and actions while driving. B) We introduce the use of range cameras as an embedded intelligent sensor [29] for monitoring the driver's body activity. C) A basic framework is presented for acquiring, segmenting, analyzing, and visualizing the recorded sequences of depth frames. Furthermore, we present a novel method for tracking the driver's motion in real time by fitting a 7-point upper body sekeltal model to the depth frames. Finally, we demonstrate the efficacy of the proposed methods using several experimental results from a pilot study.

II. Methods

Each data frame captured by a digital range camera is a two dimensional array of depth values (i.e., distance between the plane of the sensor and the depicted 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 t 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 Fig. 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. For example, Fig. 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 are shown in Fig. 1.

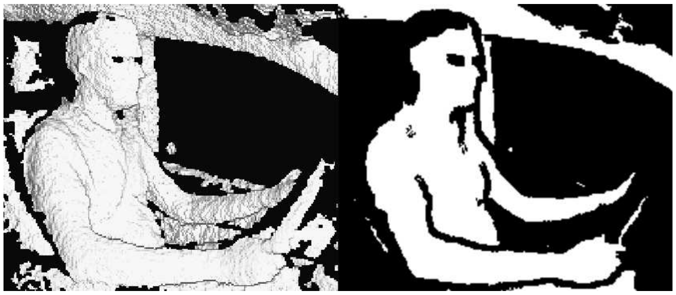

Fig. 2. Left: visualization of a depth frame. Right: The corresponding mask with enhanced boundaries between objects, computed using our framework.

The segmentation of the depth frames is a necessary pre-processing step for analyzing the activities of the human body. The process of image segmentation is a well-studied computer vision problem [30], which may be inaccurate when adjacent regions have similar color patterns, and there is no clear boundary between them. For an in-depth presentation and comparison of image segmentation algorithms the reader is referred to [30], which dedicates a chapter in mid-vision problems including segmentation. 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 computed by evaluating the following two conditions for every pixel x, y and frame t

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

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, while 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. Fig. 2 shows an example of a computed mask with clear outlines around the depicted objects.

The masked depth frames are feeded as input to a graph-based skeleton fitting algorithm that traces key body features, which is the primary goal of our data processing method. The body features of our interest include the X, Y, Z coordinates of the wrists, elbows, and shoulders as well as the orientation of the driver's torso. The values of these quantities can be estimated by fitting a human skeletal model to each of the depth frames in our datasets. The main challenge in the skeletal fitting process is that the human body in our particular field of view is very close to other objects such as the driver's seat, the steering wheel and the driver's door. Any generic skeletal fitting algorithm performs better when the human body is clearly visible and at a distance from nearby objects [23], and therefore will fail in our in-cabin setting. For instance, the skeleton tracking algorithm included in the Microsoft Kinect Software Development Kit (SDK) fails in detecting the driver's body (see Fig. 3), which motivates the development and use of a special-purpose tracking algorithm for in-cabin environments.

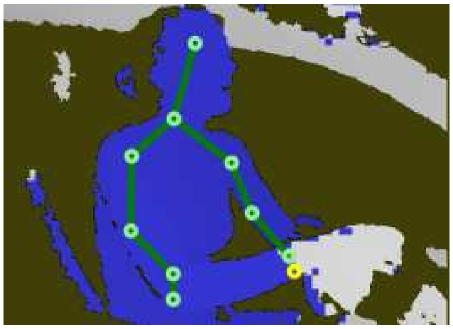

Fig. 3. Example of the skeleton model that was erroneously fit to an arbitrary frame of the depth sequence by the skeleton tracker provided with the Microsoft Kinect SDK.

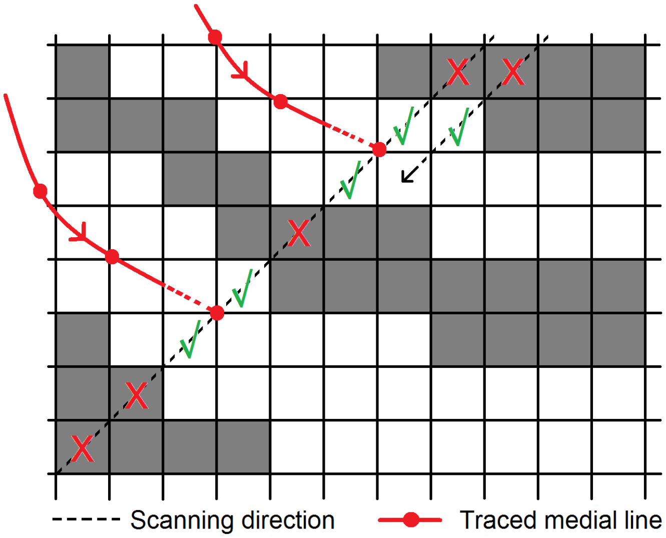

Fig. 4. Illustration of the skeleton tracking algorithm. The pixels of the masked depth frames are scanned diagonally and the medial lines of the body regions are traced (shown in red) using a graph-based algorithm. The medial curves are then filtered in order to form the driver's skeleton.

In order to overcome the aforementioned skeleton fitting challenges we developed a novel graph-based algorithm that was designed to fit a 7-point skeletal model to the body of the driver using a sequence of depth frames. Our skeletal model included the line segments between the following joints: right wrist, right elbow, right shoulder, neck, left shoulder, left elbow, and left wrist. The skeletal model is visualized in the Experimental Results section in Figs. 12 and 13. In our visualization we also show the triangle formed by the left shoulder, the right shoulder and the neck, whose normal vector was used as an indicator of the torso orientation.

The proposed skeleton fitting algorithm scans the depth frames in a diagonal fashion from upper right to lower left (see illustration in Fig. 4), pixel strip by pixel strip until the entire image is covered and segmented into line strips that are smoothly-varing 1-pixel-wide regions defined as

$$\mathcal{L} = \{(i_s, j-i_s), \cdots, (i_e, j-i_e) : i_s < i_e, \quad (1)$$
$$\left|\frac{\partial D_{i,j-i,t}}{\partial i}\right| < \epsilon_1, \left|\frac{\partial^2 D_{i,j-i,t}}{\partial i^2}\right| < \epsilon_2 \ \forall i \in (i_s, i_e)\}$$

where i_s and i_e denote the start and end pixel coordinates of the line segment, which lies on the line strip $(i, j-i)$. The length of a line segment can be easily computed by $length(\mathcal{L}) = \sqrt{2}(i_e - i_s + 1)$.

The computed line segments are organized in the form of a directed graph, which is constructed simultaneously with the segmentation of the line segments. In such graph each line segment $\mathcal{L}$ can be connected with line segments in the previous row of pixels that form the set of $parents(\mathcal{L})$ defined as

$$\mathcal{L}' \in parents(\mathcal{L}) \Leftrightarrow \exists (i, j-i) \in \mathcal{L}, \exists (i, j-i-1) \in \mathcal{L}' \quad (2)$$
$$: \left|\frac{\partial D_{i,j-i,t}}{\partial j}\right| < \epsilon_1.$$

Equivalently, each line segment can be connected with line segments in the next row of pixels by defining the set $children(\mathcal{L})$ as the inverse of Eq. 2 as follows:

$$\mathcal{L}' \in children(\mathcal{L}) \Leftrightarrow \mathcal{L} \in parents(\mathcal{L}'). \quad (3)$$

The graph produced by Eqs. 2 and 3 may contain cycles. To enforce the creation of non-cyclic graphs we define the set $father(\mathcal{L})$ as the subset of $parents(\mathcal{L})$ that contains the largest line segment:

$$father(\mathcal{L}) = \underset{\mathcal{L}' \in parents(\mathcal{L})}{\arg\max} \ length(\mathcal{L}'). \quad (4)$$

The above process segments a given depth frame into several regions that are computed as independent disconnected graphs and typically correspond to different objects in the field of view. In most applications the subject of interest corresponds to the graph with the largest number of pixels, and in general can be easily isolated from the rest of the objects in the scene.

Each graph can be further segmented into smoothly varying regions by constructing sets of connected line segments with coherent structural characteristics as follows:

$$\mathcal{S} = \{\mathcal{L}_1, \cdots, \mathcal{L}_n : \mathcal{L}_i = father(\mathcal{L}_{i+1}), \quad (5)$$
$$|children(\mathcal{L}_i)| = 1 \ \forall i \in [i, n-1]\}.$$

The line segments $\mathcal{L}_i$ in Eq. 5 form a sequence of descendants without siblings, which corresponds to a linear graph. The set of segments $\mathcal{S}$ can also be organized into a graph by defining the $father(\mathcal{S})$ and $children(\mathcal{S})$ using the connections defined in $father(\mathcal{L}_1)$ and $children(\mathcal{L}_n)$ respectively.

In our application, the regions of the arms of the depicted subjects can be found by performing simple graph searches. More specifically, the arms can be detected by searching for the two longest ancestor-child paths in the constucted graph with a common ancestor. The medial line curves of the corresponding segments along these two paths are shown in red in Fig. 4. It should be noted that the medial curves are calculated in 3D and not in the 2D coordinates of the frames. After that, the detected medial curves are filtered with an 1-dimensional Gaussian filter so that potential noise caused by the depth sensor is removed. Finally, the points that correspond to the elbows, wrists, and shoulders are estimated using spatial constraints as well as geometrical constraints regarding the size, orientation and curvature of the arms. More specifically, the elbows are estimated as the points that belong to the medial lines and have the largest distance from the line segment formed by the end points of each medial line. Similarly, the location of the wrists and shoulders are estimated in relationship to the location of the corresponding elbow.

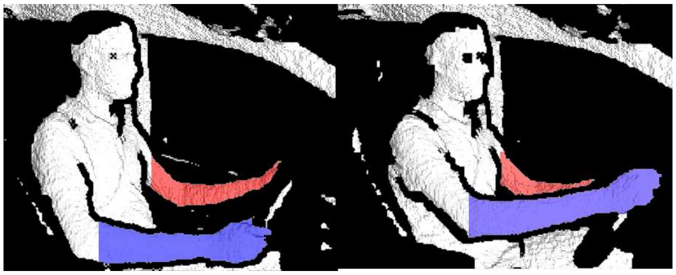

Fig. 5. Two examples of the proposed arm segmentation. Both arms can be clearly segmented from the rest of the depth frame even when one arm is occluded or partially visible from the depth camera (right).

This process fits our 7-point skeletal model to the best matching medial curves. This graph-based algorithm has linear complexity, which allow us to perform the fitting of the skeleton in real time in less than 15 milliseconds per depth frame in the computer configuration described in section III.

After fitting the skeletal model to each depth frame, we used the location of the traced joints in order to segment the original mask into regions that correspond to the arms, forearms, head, and torso using the algorithm described in [22], and the average X, Y, Z coordinates were computed from the pixels of each region. Fig. 5 shows two examples of arm segmentation. By observing the images, it is evident that the arms were accurately segmented independently of the relative position of the two arms.

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 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. 7, 8, 9).

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 Fig. 7).

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. A schematic of the study site is presented in Fig. 6. The arterial segment has three through lanes per direction, several median openings, and includes a total of six signalized intersections. 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^{TM}depth sensor contained in the Microsoft Kinect^{TM}device. The device was connected (via a USB 2.0 port) to a 64-bit computer with Intel Core i5 (quad core) 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 0.4m 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. Fig. 1 shows the field of view of the depth camera. The green rectangle depicts the closest plane of sensing, which is located 0.4m in front of the sensor (shown as the tip of the yellow pyramid in Fig. 1).

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 [22] and is available at http://research.dwi.ufl.edu/ufdw/j4k.

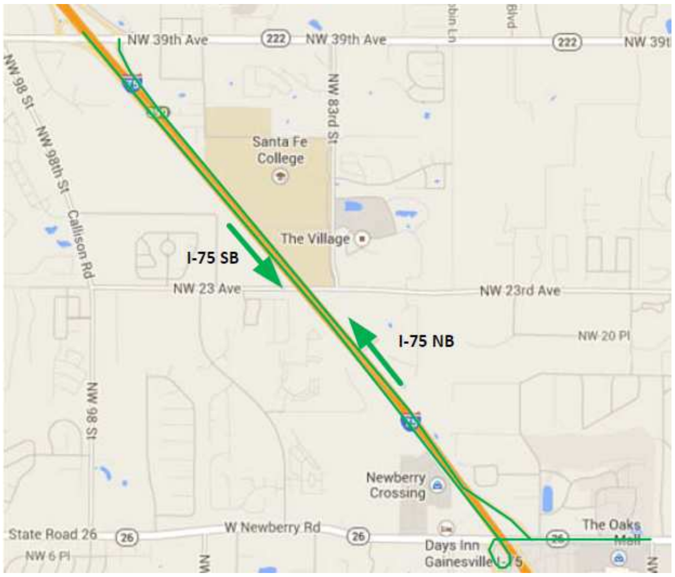

Fig. 6. Map of the route on the interstate I-75 followed in this pilot study.

Fig. 7 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. 7. 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. 7 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. 7 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, Figs. 8 and 9 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.

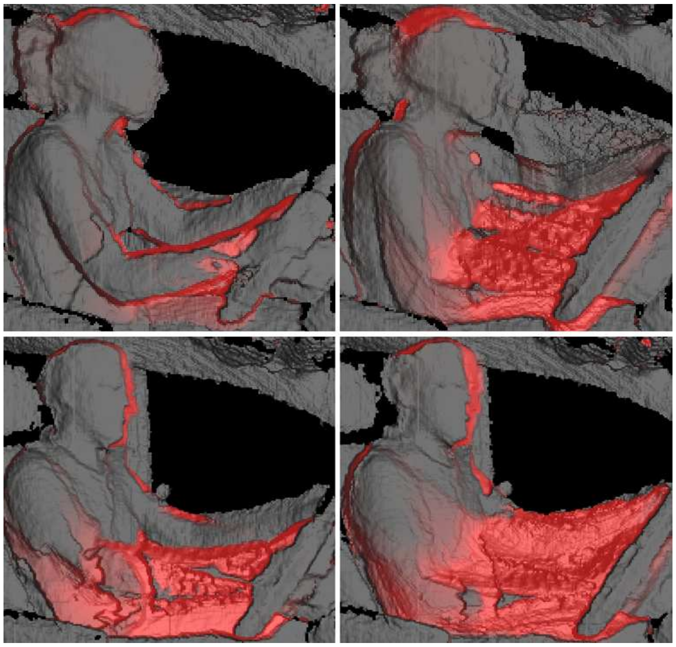

Fig. 7. Global statistics (mean and standard deviation across time) computed during merging (left) and exiting (right) from a highway for two different drivers (upper and lower row respectively). The mean is shown as the depth value within each pixel, and the st. dev. is shown as the intensity of red.

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. 10 and 11 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 changes are apparent in these figures. We further note the variability observed due to driver behavioral attributes and also due to traffic conditions. For instance, Drivers SB#2 and SB#4 appear to have increased arm and head movement compared to Drivers SB#1 and SB#3. In addition, Driver SB#2 appears to have increased head activity at several instances (e.g., note the three peaks in the graph of Fig. 10), 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.

Furthermore, using the fitted skeletal model we examined differences in the body posture during a lane change maneuver for two of the four participants. Fig. 12 shows the seven point skeleton model before and after a lane change maneuver for Driver SB#1 and Driver SB#4. The differences in the body posture between the two drivers are apparent from this figure.

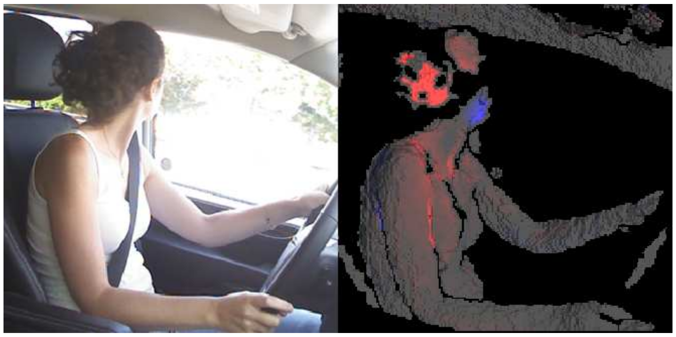

Fig. 8. An example of a video frame detecting intense head motion. On the right the corresponding computed directional magnitude of motion is shown in red or blue for increase or decrease of the depth values.

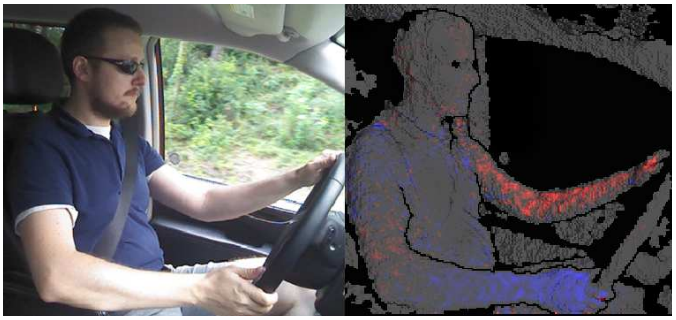

Fig. 9. Another example with intense motion of the arms. The colors and images are presented using the same format as in Fig. 8.

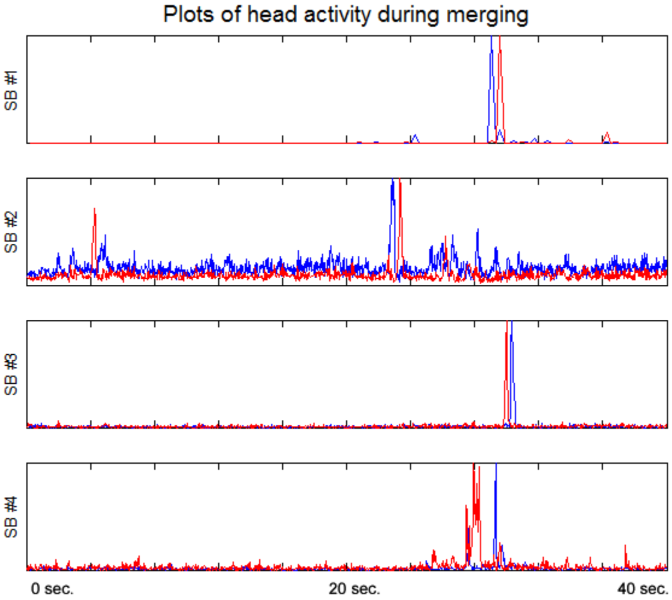

Fig. 10. Plot of the directional magnitude of the head motion during merging for 4 different drivers. The red and blue colors represent increase and decrease of the depth values respectively.

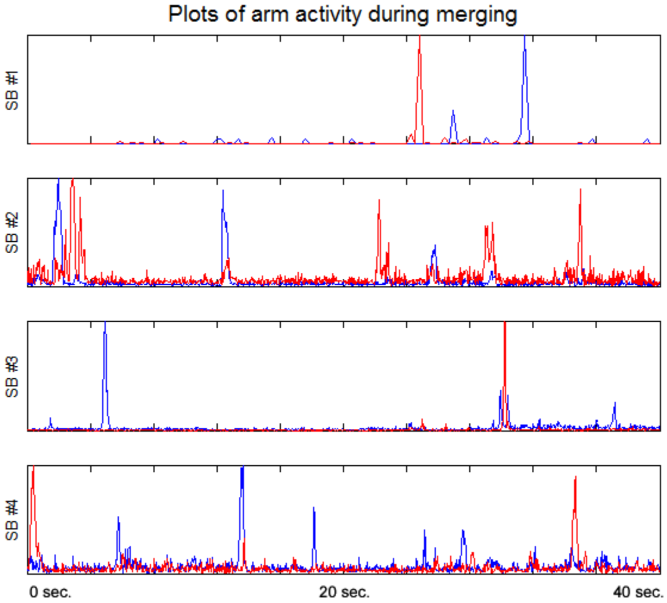

Fig. 11. Plot of the directional magnitude of the arm motion during merging. The format of the plots is the same as in Fig. 10.

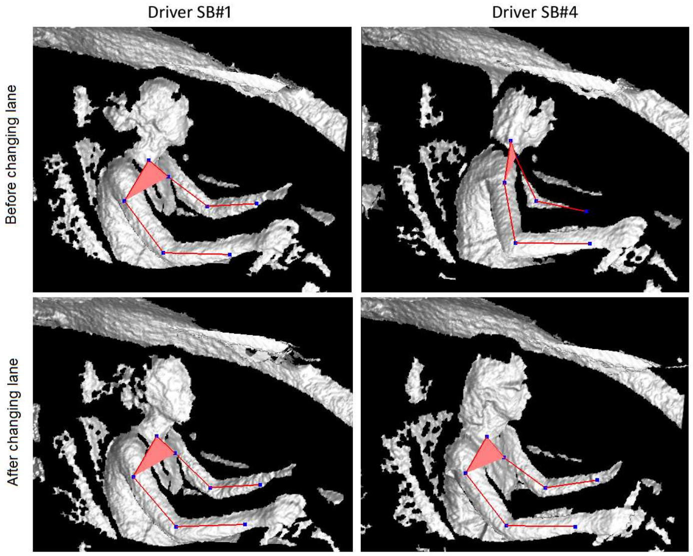

Fig. 12. Change in body posture due to a lane change maneuver shown by the fitted skeletons in the depth frames.

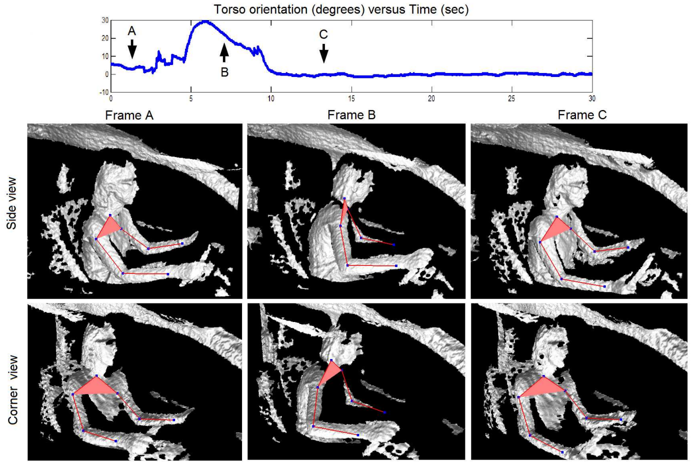

Fig. 13. Change in body posture due to a merging maneuver for Driver SB#4. The plot shows the torso orientation during the merging maneuver. The depth frames and fitted skeletons of 3 frames are shown from two 3D perspectives.

The torso of Driver SB#1 remains practically unaltered during the maneuver, whereas Driver SB#4 clearly shifts her body to the left in order to have a better visual of the traffic at the next lane. On the other hand, Driver SB#1 shifts only the head to identify potential conflict at the next lane through the rear mirror.

In addition to the lane change maneuver, a comparative analysis of the body posture during a merging maneuver was performed. Fig. 13 presents the frame sequence during a merging maneuver for Driver SB#4, along with the corresponding time-series of the torso orientation. In this graph the torso orientation represents the rotation in degrees from the torso position perpendicular to the steering wheel. The orientation is positive for left-turn rotation and negative for right-turn torso rotation. Frames A, B and C are taken as before, during, and after the execution of the merging maneuver. From these graphs it is clear that the torso rotation of Driver SB#4 is considerably increased during the merging task. Driver SB#4 torso orientation during this merging maneuver is consistent with the lane changing example shown in Fig. 12.

Similarly to the torso orientation, a comparative analysis of other parts of the participants body motion can be performed. Fig. 14 shows the time-series of the X, Y, Z coordinates of the wrists, elbows, shoulders for Driver SB#4, during the entire duration of the driving task. Using the data shown in Fig. 14 it is easy to obtain instances where there is significant body activity by identifying spikes in the respective graphs, and further analyze the underlying conditions for these instances.

By observing Fig. 14 it is evident that there is a more frequent arm motion detected during the arterial segments compared to the freeway segments as we anticipated. For example in this dataset the driver started merging onto the freeway at 100 sec. and exited at 300 sec. which correspond to intense arm activity as indicated by a significant change to the coordinates of the wrists and the elbows. During the freeway segment (between 100 sec. and 300 sec.) there was no significant change of posture detected and the coordinates of the traced joints changed only occasionally as it was also anticipated. This smooth driving pattern is significantly different compared to the one observed during the arterial segments which corresponds to 0 sec. - 100 sec., 300 sec. - 480 sec., and 650 sec. - 700 sec. During these segments the driver stopped at red traffic lights and followed a path that included many 90-degree turns. All of these instances were naturally associated with arm activity, which corresponds to changes in the coordinates of the wrists slightly as it is shown in Fig. 14.

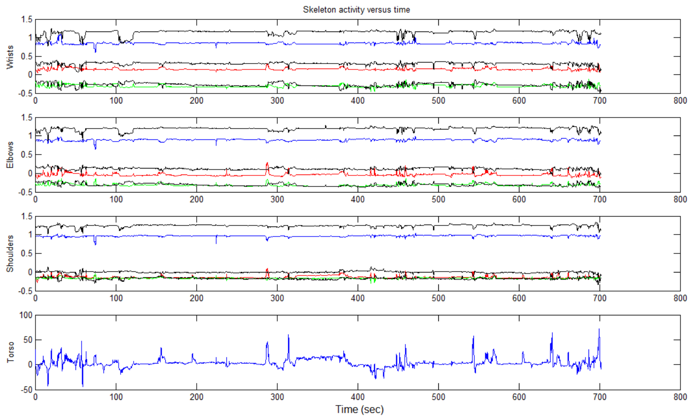

Fig. 14. Skeleton activity versus time for Driver SB#4.

Finally, the segment from 480 sec. - 650 sec. corresponds to the northbound freeway segment, which was associated with occasional body motion according to the plots in Fig. 14.

It should be noted that the proposed analysis 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, the directional magnitude of motion, and the position of the traced joints in our skeletal model for different driving maneuvers. These findings are useful for 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: Limitations and future directions

A novel approach for assessing drivers' body movements inside a vehicle was introduced in this paper. The proposed method can be used in the future for investigating how different driver types perform various maneuvers and which specific movements are associated with safe or unsafe driving conditions. This framework intends to fill a gap in literature as discussed in Sec. I and offers a tool for studying unexplored research questions regarding the correlation of drivers' body motion with potentially unsafe driving conditions.

Such research questions were intentionally not explored in this paper due to the limited number of participants in our pilot study, which is an obvious limitation of this work. The 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, and the proposed 7-point skeleton tracking revealed significant differences for different maneuver types and among the participants. Contrary to earlier research, the proposed methodology may be used for studying the upper body posture and motion as a whole, instead of focusing on individual parts of the body in isolation, overcoming the limitations of general purpose tracking algorithms. However, the main strength of our method can be also cosidered a limitation since our algorithm is optimized for tracking human activities in a vehicle's cabin environment and is not suitable for general use such as in gaming.

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. 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 / DTRT12-G-UTC04 US Department of Transportation / FDOT / ALDOT. The authors would like to thank the sponsors and the anonymous volunteers who participated in the 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," *IEEE Trans. on Intelligent Transportation Systems*, 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,” *IEEE Trans. on Intelligent Transportation Systems*, 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] 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.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] A. Barmpoutis, E. Fox, I. Elsner, and S. Flynn, “Augmented-reality environment for locomotor training in children with neurological injuries,” *Proceedings of the Workshop on Augmented Environments for Computed Assisted Interventions, LNCS*, vol. 8678, pp. 108–117, 2014.

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

[25] B. Lange and et al., “Interactive game-based rehabilitation using the Microsoft Kinect,” *IEEE Virtual Reality Workshops*, pp. 171–172, 2012.

[26] I. Oikonomidis and et al., “Efficient model-based 3d tracking of hand articulations using Kinect,” *In Proc. of the British Machine Vision Association Conference*, 2011.

[27] S. Adamovich, G. Fluet, E. Tunik, and A. Merians, “Sensorimotor training in virtual reality: a review.” *NeuroRehabilitation*, vol. 25, no. 1, pp. 29–44, 2009.

[28] A. Kondyli, V. Sisiopiku, and A. Barmpoutis, “A 3d experimental framework for exploring drivers’ body activity using infrared depth sensors,” *IEEE International Conference on Connected Vehicles*, pp. 574–579, 2013.

[29] B. Guo, D. Zhang, Z. Yu, Y. Liang, Z. Wang, and X. Zhou, “From the internet of things to embedded intelligence,” *World Wide Web*, vol. 16, no. 4, pp. 399–420, 2013. [Online]. Available: http://dx.doi.org/10.1007/s11280-012-0188-y

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

Alexanrda Kondyli received the B.Sc. degree in rural and surveying engineering from the National Technical University of Athens, Athens, Greece, in 2003, the M.Sc. and Ph.D. degrees in civil engineering (transportation) from the University of Florida, Gainesville, USA, in 2005, and 2009 respectively. Currently, she is an Assistant Professor in the Department of Civil, Environmental & Architectural Engineering at the University of Kansas. Her research interests include driver behavior, traffic operations, traffic safety and microsimulation.

Virginia Sisiopiku is an Associate Professor in Civil, Construction, and Environmental Engineering at the University of Alabama at Birmingham (UAB) and the UAB Transportation Program Director. She has over 20 years of transportation experience in the areas of intelligent transportation systems, traffic operations and safety, and simulation. She is the recipient of the 2010 Deans Award for Excellence in Mentorship, the 2007 Presidents Excellence in Teaching Award and a Fellow of the Institute of Transportation Engineers (ITE).

Liangke Zhao received the B.Sc. degree in computer science and technology from the Capital Normal University, Beijing, China, in 2012. He is currently pursuing his M.Sc. degree in computer engineering at the University of Florida, Gainesville, USA. His research interests lie in the areas of computer vision and copmuter graphics.

Angelos Barmpoutis received the B.Sc. degree in computer science from the Aristotle University of Thessaloniki, Thessaloniki, Greece, in 2003, the M.Sc. degree in electronics and electrical engineering from the University of Glasgow, Glasgow, U.K, in 2004, and the Ph.D. degree in computer and information science and engineering from the University of Florida, Gainesville, USA, in 2009. Currently, he is an Associate Professor and the coordinator of research and technology in the Digital Worlds Institute, University of Florida. Dr. Barmpoutis' research interests lie in the areas of 3D machine vision, biomedical imaging, digital humanities, and human-computer interaction.