DHM2020
L. Hanson et al. (Eds.)

doi:10.3233/ATDE200024

A Pipeline for Creating In-Vehicle Posture Database for Developing Driver Posture Monitoring Systems

Mingming ZHAO [a,b], Georges BEURIER [a], Hongyan WANG [b] and Xuguang WANG [a,1]
[a] *Univ Lyon, Univ Gustave Eiffel, LBMC UMR_T9406, F69622, Lyon, France*
[b] *Tongji University, School of Automotive Studies, 200092, Shanghai, China*

Abstract. Driver posture monitoring is beneficial for identifying driver physical state as well as for optimizing passive safety systems to mitigate injury outcomes during collisions. In recent years, depth cameras are increasingly used to monitor driver's posture. However, good driver posture data is missing for developing accurate posture recognition methods. In this study, we introduce a method to build an in-vehicle driver posture database for training posture recognition algorithms based on a depth camera. Driver motion data was collected from 23 participants performing both driving and non-driving activities by an optical motion capture system Vicon. Motions were reconstructed by creating personalized digital human skeletons and applying inverse kinematics approach. By taking advantage of the techniques developed in computer graphics, a recorded driver motion can be efficiently retargeted to a variety of virtual humans to build a large database including synthetic depth images, ground truth labels of body segments and skeletal joint centers. Examples from motion reconstruction, data augmentation and preliminary posture prediction results are given.

Keywords. Driver Posture, Motion Reconstruction and Animation, Motion Database, Driver Monitoring

1. Introduction

According to the global status report on road safety released by World Health Organization in 2018, road traffic accidents claimed over 1.35 million fatalities and more than 50 million serious injuries worldwide each year. In spite of the advances in vehicle safety technologies, most of these accidents are still caused by human driver errors [1] and therefore could be avoidable. Driver monitoring has been an important research topic for many years. Especially with the background of driving automation of level 3 (SAE standard), researchers are struggling to find efficient ways to assess driver state and develop driving assistance systems that can deal with the human-machine transition in a smooth and safe manner [2][3]. Driver posture is an important source of information to evaluate driver's state, such as distraction [4] and inappropriate operation [5]. Furthermore, dynamic tracking of driver posture is necessary for developing future intelligent restraint systems that can be automatically adapted to driver's position to prevent or reduce injury during collision [6].

[1] Corresponding Author, Email: xuguang.wang@univ-eiffel.fr.

Recently, thanks to the progress in 3D modeling and depth perception in computer vision community, several non-invasive measuring methods for human posture recognition have been proposed [7]. As opposed to conventional cameras, depth cameras are not subject to varying illumination conditions, color, and texture in the scene and most importantly demonstrate a good compromise between real-time performance and 3D measurement [8]. Thanks to these characteristics, depth cameras have been introduced to monitor driver's upper body posture [2][4][9]. In these studies, Microsoft Kinect camera was adopted along with the integrated real-time posture recognition algorithm predicting 3D joint positions from a single depth image [10]. This algorithm was featured by the classification of body parts using Random Forest classifier [11] followed by the estimation of 3D joint positions using Meanshift [12]. Millions of human pose depth images with ground truth labels were artificially generated to train the algorithm. However, the algorithm was mainly intended for gaming. Training samples were collected from entertainment activities, such as dancing, fighting, etc. When directly used to monitor drivers, posture recognition suffers due to body occlusions and suboptimal camera placement in the vehicle cabin [13]. Therefore, there is a need to develop more performant in-vehicle postural monitoring algorithm based on depth images.

A reliable posture recognition algorithm requires a large amount of data samples for training and evaluation to avoid overfitting. To facilitate the research of driver posture recognition using depth cameras, more and more in-vehicle driver posture databases including depth images are now publicly available [14][15][16]. The database in [16] even provided drivers' ground truth motion capture data for reference. One crucial information still missing in these databases is the body segmentation label. To annotate on depth images of a real driver is challenging. To this end, Yamada et al. [17] dressed two drivers with a color-coded T-shirt to generate labeled depth images in order to customize the existing posture recognition algorithm [10]. However, the ground truth joint centers were not collected in the experiment. Apart from that, this method for generating body part labels is time consuming and the T-shirt only covered the trunk and arms.

The aim of this work was to establish a pipeline for creating in-vehicle driver posture database necessary for developing driver posture monitoring systems based on a depth camera. Our goals for building this pipeline are two-fold. On one hand, the data samples generated by this pipeline contain a good coverage of realistic posture variations, more specifically the appearance variations in depth images. On the other hand, the ground truth labels including body part segmentation and joint centers should be accurately obtained in an efficient way.

2. Materials and method

Figure 1 presents an overview of the pipeline, which mainly consists of three parts: data collection, motion reconstruction and data augmentation. We first perform experiments to collect a large set of driver posture data using both a motion capture (mocap) system VICON and a Microsoft Kinect. Then, driver motion is reconstructed from the raw mocap data using RPx [18], a customized human model based motion analysis and simulation tool. A personalized skeleton template is defined and inverse kinematics is applied to calculate the orientation and position of each body segment during movement.

Although depth images are also collected from the experiment and ground truth joint centers are estimated, it is not a good idea to directly use these samples to train the posture recognition algorithm due to the complexity in manual body part labeling, because this is tedious and cumbersome at best and potentially could lead to label inaccuracy. In addition, participants in an experiment are usually dressed with tight gym suits. Thereby depth images obtained in lab conditions are clearly different from those in real conditions.

Inspired by the work in [10], we resort to data augmentation by means of computer graphics techniques to obtain a large number of synthetic depth images covering a wide variety of driver postures and more importantly to simplify the collection of the ground truth labels. The basic idea is to animate rigged virtual human characters with realistic external envelopes including skin, clothes, hair and predefined body part labels using reconstructed motions captured in lab condition. This allows us to synthesize artificial depth images by rendering and meanwhile provides ground truth labels including body part segmentation and joint centers. The virtual human characters are created in MakeHuman (MH) [19], an open source tool designed to prototype realistic 3D human models. The motion retargeting, image rendering and data recording process are performed in MAYA [20], a 3D computer animation, modeling, simulation, and rendering software.

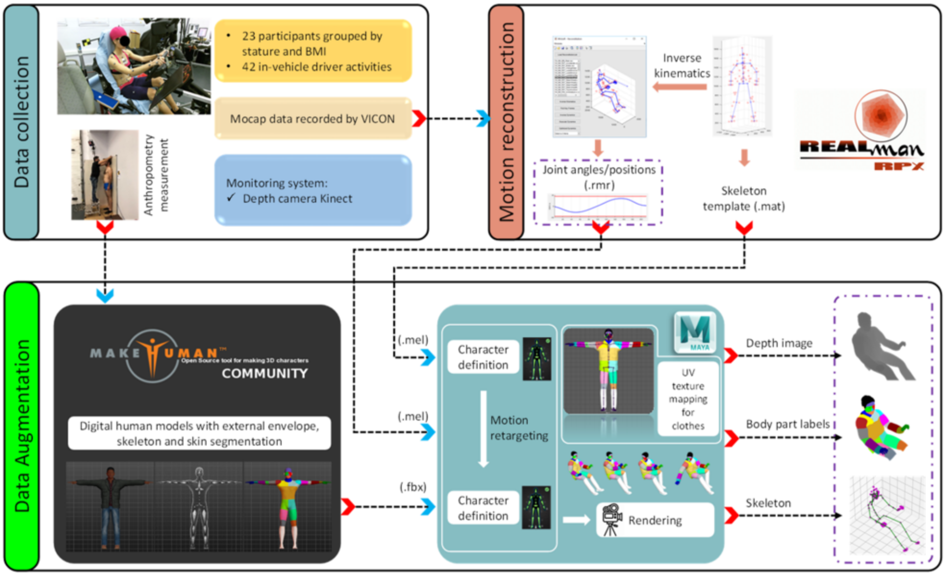


Figure 1. Overview of proposed pipeline.

2.1. Driver motion data collection

Twenty-three drivers (12 males and 11 females) with driving experience, ranged in age from 22 to 65 years, in height from 153 cm to 195 cm, in Body Mass Index (BMI) from 18.2 kg/m^2 to 43.4 kg/m^2, took part in this study. All participants provided written informed consent prior to participation in this study. The experimental protocol was approved by the ethical committee of Gustave Eiffel University (formerly known as French Institute of Science and Technology for Transport, Development and Networks).

We captured motions corresponding to 42 representative in-vehicle driver activities with reference to the work in [21]. They included primary driving operations, such as switching gear, turning steering wheel, etc., common secondary behaviors, such as interacting with phone, reading the dashboard, etc., and non-driving activities that might be present in autonomous vehicles, such as holding a book with both hands, relaxing both feet on the floor, etc. During experiment, participants were instructed to perform these tasks one by one on an experimental mockup (Figure 2a).

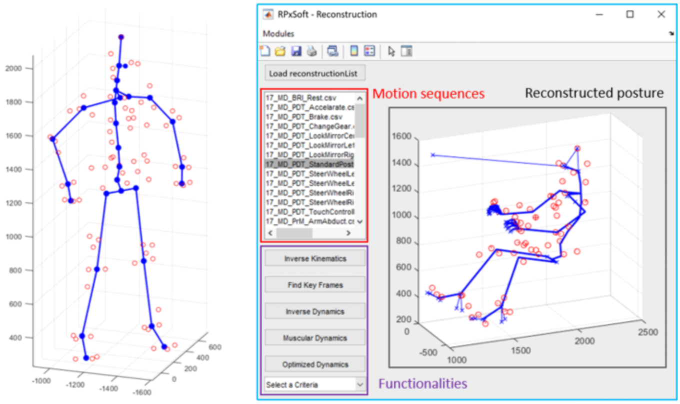
A motion capture system Vicon Giganet® MXT (Vicon Motion Systems, Oxford, United Kingdom) with 14 infrared cameras was used to track reflective markers placed on participants (Figure 2b) at a frequency of 50 Hz. A Microsoft Kinect camera was placed right front of the driver to observe upper body posture through depth perception with a frequency of 25 Hz. The synchronization between two systems was realized by an electronic trigger.

(a) (b)

Figure 2. Experimental mockup (b) and marker arrangement (a).

2.2. Motion reconstruction

From measured marker trajectories, body motion was reconstructed using RPx. First, a subject specific articulated skeleton was created for each participant from a reference standing posture (Figure 3a). The joint positions related to hips, spine, shoulders and neck were estimated by statistical models [22][23]. Regarding limbs and head, joint positions were simply determined as the center of the marker pair attached close to the target joint or body part. Once the personalized skeleton template was created, joint angles were calculated by minimizing the distance between model based marker positions and measured ones (Figure 3b). Note that the real hand gestures of the participants were not collected in detail. The default hand gesture from RPx was adopted for all participants. Refer to [24][25][26] for more details. Motion reconstruction process was implemented in Matlab.



(a) (b)

Figure 3. Personalized skeleton template (a) and RPx motion reconstruction module (b).

2.3. MH human character

The basic human character in MakeHuman was composed of a rigged skeleton and a skin mesh that forms the body surface shape. The skeleton (Figure 4 left) was linked to the mesh (Figure 4 middle) using the linear blend skinning technique available in the Blender software [27] so that changes of joint positions led to an adaptation of skin mesh. To impersonate the real participants, the attributes for each virtual model (e.g., age, gender, weight, height, body proportion etc.) were adjusted according to subject's anthropometry measurements. We also randomly configured clothes and hairstyle materials for each base character to yield more realistic driver appearances in depth images. We defined 30 color-coded body parts that densely cover the body, as suggested by [10]. These parts were specified in a texture map that can be applied to the body skin of different models (Figure 4 right). The resulting models were exported in standard FBX files that can be seamlessly manipulated in MAYA.

Figure 4. Digital human models for each participant. Skeleton (left), body shape with clothes and hair (middle) and body shape with segmentation labels on the skin (right).

2.4. Synthetic data generation

The reconstructed motion data including the RPx skeleton template and joint angles were recompiled using mel (the scripting language of MAYA) so that they were compatible with the animation engine in MAYA. With help of Autodesk® HumanIK® (HIK) animation middleware (a full-body inverse kinematics (IK) solver and retargeter), the motion of RPx skeleton could be easily retargeted to MH skeleton.

Since the clothes would shadow the colors attached to the body skin during rendering, we mapped the same texture onto various clothes in MAYA UV editor as to the body skin in MH, so that the body segmentation labels were consistent across different characters dressed with different clothes (Figure 5).

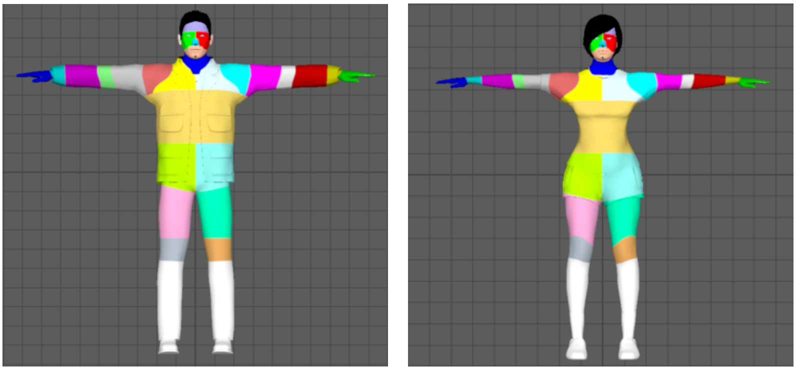


Figure 5. Body segmentation labels from skin and clothes.

During animation, the scene was rendered into color and depth images by using Arnold Renderer for MAYA (an advanced Monte Carlo ray tracing renderer) and the joint centers were recorded. In addition, we marked some key points on the head with the expectation that these points could be useful for monitoring head orientation. The motion retargeting, image rendering and data recording in MAYA were automatized by a hybrid programming of Python and mel.

3. Results

The motion database consists of approximately 2.4 Million frames in the driver motion sequences. Figure 6 gives one example after motion reconstruction. The reconstructed postures were from a motion when the driver moved the left foot and right hand to switch gear. The joint angles of left hip (GHUL) and the positions of left foot (GFBL) during motion are plotted.

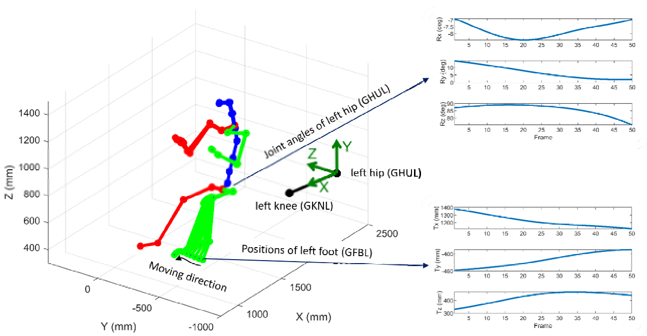


Figure 6. Motion reconstruction.

Figure 7 shows an example where we retargeted the real driver motion when reaching the phone with right hand onto a MH character. The generated depth images, body segmentation color labels and skeletal joint centers are given.

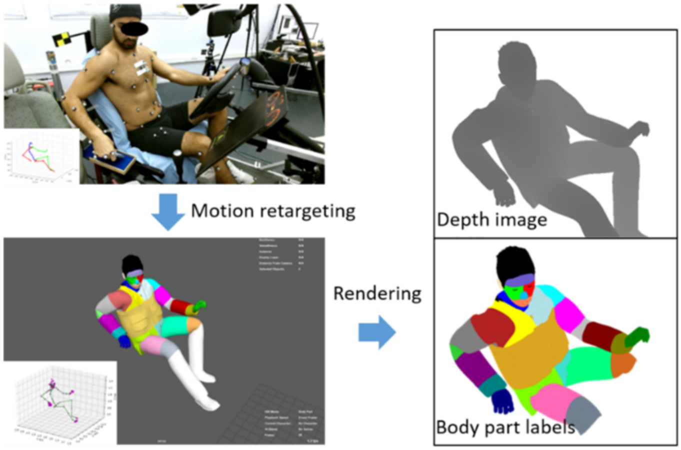


Figure 7. Data augmentation.

Given the synthetic dataset generated by the proposed pipeline, we adapted the algorithm from [10] and extracted some key joint centers and key points for prediction. As mentioned before, this algorithm transformes the body segmentation as a per-pixel classification task. After identifying the relevant body parts, the joint centers beneath the mesh surface and key points on the head could be directly inferred. Assume that the background is subtracted, our algorithm could discern different body parts and predict driver's posture from a depth image and an example is shown in Figure 8. Note that the lower body parts are merged into a single part annotated by white color and fall out of our interest, because they are practically invisible by a single camera in a real car.

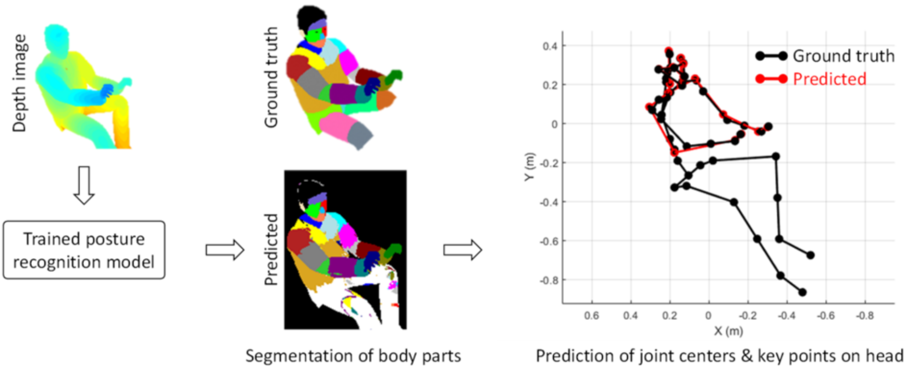


Figure 8. An example of posture recognition.

4. Discussion and conclusion

The aim of this paper is to introduce a pipeline to generate diver posture depth images and high quality ground truth labels required by the development of in-vehicle posture recognition algorithms. The pipeline is built by taking advantage of existing techniques including RPx, MakeHuman and MAYA. Results show that real driver motions can be reliably reconstructed even when some markers were missing and synthetic dataset including realistic depth images and accurate ground truth labels can be generated in an efficient manner. Compared to the publicly existing driver posture databases [14][15][16], the dataset generated by this pipeline contains more information including complete ground truth labels for depth images. In contrast to the T-shirt method used in [17], the data augmentation process proposed in the present work allows us to obtain the body part labels more efficiently.

It should be noted that the motion retargeting in MAYA is not limited inbetween the MakeHuman character and mocap character of the same participant, although this was performed in this work. Theoretically and technically, motion retargeting can be performed across different participants. By generating more characters in MakeHuman, the dataset can be potentially enlarged to an infinitive scale based on the limited mocap data.

Another advantage of data augmentation is that the virtual environment provides us the opportunity to find the optimum depth camera position in the vehicle cabin, as proposed by Plantard et al. [28].

The main limitation of the proposed pipeline lies in the nature of marker-based optical motion capture system. In order to reduce relative movement of markers with respect to body segment, participants were required to dress with tight gym suits and markers were directly attached to the skin. In addition, the seatbelt was not used. In the motion retargeting and animation process, the clothes deformed in accordance with the skeletal posture of the character but not vice versa. Therefore, possible effects of clothing were not considered in the present study.

In addition, realistic driving operations were not imposed on the experimental mockup. Although we designed 42 different trials for participants to simulate the real driver motions in various scenarios, this list was by no means exhaustive. In fact, to cover all the driver motions in realistic driving conditions is not possible and also not our primary goal. According to [10], a wide range of posture combinations of different body parts would be sufficient for generalizing the prediction model to unseen postures.

More research will be performed to further improve the data augmentation process including the segmentation layout of body parts, the integration of driver hand gesture details, etc. In addition, the driver posture recognition algorithm trained on the synthetic dataset will be tested on the real depth images collected from the experiment.

Acknowledgement

This study was funded by China Scholarship Council (CSC 201806260131) and partly supported by the French National Project ANR AutoConduct (ANR-16-CE22-0007). The authors would like to express sincere gratitude to Ilias Theodorakos and Richard Roussillon for their contribution during the design and implementation of the experiment as well as to the volunteers who participated in the experiments for their efforts and time.

References

[1] Singh S. Critical reasons for crashes investigated in the national motor vehicle crash causation survey. Traffic Safety Facts Crash-Stats. Report No. DOT HS 812 115. National Highway Traffic Safety Administration, Washington, DC, 2015 Feb.

[2] Shia VA, Gao Y, Vasudevan R, Campbell KD, Lin T, Borrelli F, Bajcsy R. Semiautonomous vehicular control using driver modeling. IEEE Transactions on Intelligent Transportation Systems, 2014, 15(6), 2696-2709.

[3] Okabe K, Watanuki K, Kaede K, Muramatsu K. Study on Estimation of Driver's State During Automatic Driving Using Seat Pressure. In: International Conference on Intelligent Human Systems Integration, 2018, 35-41. Springer, Cham.

[4] Xing Y, Lv C, Zhang Z, Wang H, Na X, Cao D, Velenis E, Wang FY. Identification and analysis of driver postures for in-vehicle driving activities and secondary tasks recognition. IEEE Transactions on Computational Social Systems, 2017, 5(1), 95-108.

[5] McCall JC, Trivedi MM. Driver behavior and situation aware brake assistance for intelligent vehicles, In: Proceedings of the IEEE, 2007, 95(2), 374-387.

[6] Filatov A, Scanlon JM, Bruno A, Danthurthi SS, Fisher J. Effects of Innovation in Automated Vehicles on Occupant Compartment Designs, Evaluation, and Safety: A Review of Public Marketing, Literature, and Standards. SAE Technical Paper, 2019, Apr 2.

[7] Berretti S, Daoudi M, Turaga P, Basu A. Representation, analysis, and recognition of 3D humans: A survey. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 2018, 14(1s), 1-36.

[8] Wang H, Zhao M, Beurier G, Wang X. Automobile Driver Posture Monitoring Systems: A Review. China Journal of Highway and Transport, 2019, 32(2), 1-18.

[9] Chuang YW, Kuo CH, Sun SW, Chang PC. Driver behavior recognition using recurrent neural network in multiple depth cameras environment. Electronic Imaging, 2019, (15), 56-1.

[10] Shotton J, Sharp T, Kipman A, Fitzgibbon A, Finocchio M, Blake A, Cook M, Moore R. Real-time human pose recognition in parts from single depth images. Communications of the ACM, 2013, 56(1), 116-124.

[11] Breiman L. Random forests. Machine learning, 2001, 45(1), 5-32.

[12] Comaniciu D, Meer P. Mean shift: A robust approach toward feature space analysis. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2002, (5), 603-619.

[13] Zhao M, Beurier G, Wang H, Wang X. In vehicle diver postural monitoring using a depth camera kinect. SAE Technical Paper, 2018, Apr 3.

[14] Martin M, Roitberg A, Haurilet M, Horne M, Reiß S, Voit M, Stiefelhagen R. Drive&Act: A Multi-modal Dataset for Fine-grained Driver Behavior Recognition in Autonomous Vehicles. InProceedings of the IEEE International Conference on Computer Vision, 2019, 2801-2810.

[15] Jegham I, Khalifa AB, Alouani I, Mahjoub MA. MDAD: A Multimodal and Multiview in-Vehicle Driver Action Dataset. InInternational Conference on Computer Analysis of Images and Patterns, 2019, 518-529. Springer, Cham.

[16] Feld H, Mirbach B, Katrolia J, Selim M, Wasenmüller O, Stricker D. DFKI Cabin Simulator: A Test Platform for Visual In-Cabin Monitoring Functions. arXiv preprint arXiv:2002.03749, 2020, Jan 28.

[17] Yamada T, Irie H, Sakai S. High-accuracy joint position estimation and posture detection system for driving. In: Adjunct Proceedings of the 13th International Conference on Mobile and Ubiquitous Systems: Computing Networking and Services, 2016, 219-224.

[18] Monnier G, Renard F, Chameroy A, Wang X, Trasbot J. A motion simulation approach integrated into a design engineering process. SAE Transactions, 2006, 1118-1123.

[19] http://www.makehumancommunity.org/

[20] https://www.autodesk.com/education/free-software/maya

[21] Dingus TA, Klauer SG, Neale VL, Petersen A, Lee SE, Sudweeks J, Perez MA, Hankey J, Ramsey D, Gupta S, Bucher C. The 100-car naturalistic driving study, Phase II-results of the 100-car field experiment. United States. Department of Transportation. National Highway Traffic Safety Administration, 2006.

[22] Reed MP, Manary MA, Schneider LW. Methods for measuring and representing automobile occupant posture. SAE Technical Paper, 1999, Mar 1.

[23] Peng J, Panda J, Jan SV, Wang X. Methods for determining hip and lumbosacral joint centers in a seated position from external anatomical landmarks. Journal of biomechanics, 2015, 48(2), 396-400.

[24] Monnier G, Wang X, Verriest JP, Goujon S. Simulation of complex and specific task-orientated movements-application to the automotive seat belt reaching. SAE transactions, 2003, 715-721.

[25] Wang X, Chevalot N, Monnier G, Ausejo S, Suescun Á, Celigüeta J. Validation of a model-based motion reconstruction method developed in the Realman project. SAE transactions, 2005, Jan 1, 873-879.

[26] Ausejo S, Suescun A, Celigüeta J, Wang X. Robust human motion reconstruction in the presence of missing markers and the absence of markers for some body segments. SAE Technical Paper, 2006, Jul 4.

[27] https://www.blender.org/
[28] Plantard P, Auvinet E, Pierres ASL, Multon F. Pose Estimation with a Kinect for Ergonomic Studies: Evaluation of the Accuracy Using a Virtual Mannequin. Sensors, 2015, 15(1), 1785-1803.